



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 171654

TO: Celine Qian
Location: REM-2A64/2C70
Art Unit: 1636
Monday, November 21, 2005

Case Serial Number: 10/099663

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

This Page Blank (uspio)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library Remsen Bldg.



This Page Blank (1 of 10)

A17N, Ed Hart

FOR OFFICIAL USE ONLY

171655

ACCESS DB #

PLEASE PRINT CLEARLY
RECEIVED

Scientific and Technical Information Center

NOV 16 2005

CRFG

SEARCH REQUEST FORM

CH/CHEM. DIVISION

(STIC)

Requester's Full Name: Celine Qian Examiner #: 78710 Date: 11/15/05

Art Unit: 1636 Phone Number: 2-0777 Serial Number: 101099663

Location (Bldg/Room#): 2A64 (Mailbox #): 2C70 Results Format Preferred (circle): PAPER DISK

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: Conditional K10 meth. for Gene trapping & gene targeting using
Gene inducing silencer

Inventors (please provide full names):

As Kew et al.

Earliest Priority Date: 5/30/03

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known.

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO: 1

1-2381 NA
LB

STAFF USE ONLY

Searcher: _____

Type of Search

1 NA Sequence (#)

Searcher Phone #: _____

AA Sequence (#)

Searcher Location: _____

Structure (#)

Date Searcher Picked Up: 11/16/05

Bibliographic

Date Completed: 11/21/05

Litigation

Searcher Prep & Review Time: _____

Fulltext

Online Time: _____

Other

Vendors and cost where applicable

STN Dialog

Questel/Orbit Lexis/Nexis

Westlaw W/VV/Internet

PS In-house sequence systems

☒ Commercial ☐ Oligomer ☐ Score/Length
☒ Interference ☐ SPDI ☐ Encode/Transl
Other (specify)

This Page Blank (usp10)


```
unure      complement (28639..28644)
unure      /note=<30 qual SNGL region"
unure      28734..28744
unure      /note=<30 qual SNGL region"
unure      28768..28774
unure      /note=<30 qual SNGL region"
repeat_region
repeat_region /rpt_family="GC-rich"
unure      28836..28841
unure      /note=<30 qual SNGL region"
unure      28855..28867
unure      /note=<30 qual SNGL region"
repeat_region
repeat_region /rpt_family="CA)n"

Query Match      9.0%; Score 46; DB 9; Length 163533;
Best Local Similarity 45.5%; Pred. No. 1.3;
Matches 163; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy      23 TTATTATGTCGAGTATGATATTCACCAATTGAAATGACAAAGCAGATTGTG 82
Db      48902 TTTAAACATTGATTAATAATTTATCTATAAATTTATTAACAATTAAATATG 48961

Qy      83 TTTTATCTGTGGTAATAAGTTTCTCCAGTTGTATAAAGCCCTCCACAGTATA 142
Db      48962 GATTACAGATAGATTAATTCCTGTATCCAGTGTGTTAAATGTTAAACATTCGAA 49021

Qy      143 AGTCATGACACAAAGAAATGCAATACATTCCTTACGTCATTAATTTTCATTA 202
Db      49022 AATTCATTAATTTTAAAGTAATTTATTTATTTATTTATTTATTTAGTTGAACA 49081

Qy      203 GATAGCCGTTTCTTACTACACTCAATTAAGATGAACAGAAATGAGGTAGTACTG 262
Db      49082 TTTTATTTTAAATTTTCATGCGTACATGAGTACATTAATTAATGAGGATACATGA 49141

Qy      263 TTTATAAGAGAGATATAAGATTAATCATCATCATTTGAGCAATAAGAGGAGAGAT 322
Db      49142 TATTTTGATACACACATTAATAATTCATTAATTTTAAAGATGATTTTCAAGCTTTA 49201

Qy      323 TCAGCAAAACAGTGTCTTACAGTGGAAAACAAATTAACTAAAGTGACCCCTCTCT 380
Db      49202 TTTGTATTTAGTATATAATAATGACATAGTTTAAAGTAAGTATCTCACCTCTT 49259

RESULT 2
AP003189      298050 bp      DNA      linear      BCT 20-MAR-2004
LOCUS      Clostridium perfringens str. 13 DNA, complete genome, section 5/10.
DEFINITION      AP003189 BA000016
ACCESSION      AP003189.2 GI:18146727
VERSION
KEYWORDS
SOURCE
ORGANISM
Clostridium perfringens str. 13
Clostridium perfringens str. 13
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE
AUTHORS      Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A.,
TITLE      Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S. and Hayashi, H.
JOURNAL      Complete genome sequence of Clostridium perfringens, an anaerobic
MEDLINE      flesh-eater
PUBMED      Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
AUTHORS      Hattori, M., Yamashita, A., Oshima, K. and Shiba, T.
TITLE      Direct Submission
JOURNAL      Submitted (15-FEB-2001) Masahira Hattori, Kitasato Institute for
LIFE Sciences, Kitasato University, Kitasato 1-15-1, Sagamihara,
Kanaagawa 228-8555, Japan
URL: http://genome.1s.kitasato-u.ac.jp,
E-mail: hattori@genome.1s.kitasato-u.ac.jp,
Tel: 81-42-778-8194,
Fax: 81-42-778-8193
COMMENT      On Jan 14, 2002 this sequence version replaced gi:18144662.
```

This work was done in collaboration with Tohru Shimizu, Kaori Ohtani, Hideo Hayaishi (University of Tsukuba), Hideki Hirakawa, Satoru Kuhara (Kyushu University), Naotake Ogasawara (Nara Institute of Science and Technology), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science.

FEATURES

source

```
Location/Qualifiers
1. 298050
/organism="Clostridium perfringens str. 13"
/mol_type="genomic DNA"
/strain="13"
/db_xref="taxon:195102"
/note="anaerobic pathogen for gas gangrene"
complement(34..1056)
/gene="CPE1002"
complement(34..1056)
/gene="CPE1002"
/note="340 aa, no significant homology."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80708.1"
/db_xref="GI:18144663"
/translation="MSNKKINPKNMNSDPKVLINSLVLDGISLKEISKIDIOKDD
IIRKMBEGFPDSDGDFPKDKESTSDIDPSVTSKSLPKPERPDVOSIISI
NEKLNKDITETTNILPEIKTSEKSVTKEDNEDTKTSQSTLEIPVRSYARKK
KKRKTILKMIKISYFNKSFSLIKLKKKSSSKLECNPFKNKEVINIKSTILLE
VAIEKLIPVNDIKETSHGNSSEKLPFEIKIPKPFGRIRIPVDDLDIEINFLND
VSLIEDKDFDKNSDPIKINSDSLSEBNDKSLSEBVCNLEKKINELEPFLIF
QAMSNK"
1361..1540
/gene="CPE1003"
1361..1540
/gene="CPE1003"
/note="59 aa, similar to gnu:AP001515.87 BH2353 gene
product from Bacillus halodurans (79 aa); 46.7% identity
in 45 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80709.1"
/db_xref="GI:18144664"
/translation="NAIDELTKRINELHKKHKEBGLSEBHKEREELREKYNFRNSN
LKEQLKGIPEKKNKN"
1552..2475
/gene="hprk"
1552..2475
/gene="hprk"
/note="307 aa, similar to gnu:AP001519.105 Hpr (Ser/Thr)
protein kinase/phosphatase from Bacillus halodurans (310
aa); 45% identity in 282 aa overlap
CPE1004"
/codon_start=1
/transl_table=11
/product="probable Hpr. Ser/Thr protein kinase/phosphatase"
/protein_id="BAB80710.1"
/db_xref="GI:18144665"
/translation="MGVTEKLIKDPSELEIVIGSEBNPYNVDVNPQLQAGFYNY
FPERIOVGKAEWSEFLEMDSPDLKRLNKPSPDISCLITRGLIEHLELKARK
RLMLTIRSMVMTTKTISKITMYLSDMKAPETPLHGVLDVNGMLTIGESGKSET
ALELIRKGRHLYTDDAVDIKEIDGLIGSPETITGMLLEVRSKGIIDVSAIYGLSSTL
NSKQIKIILHFEHMKDDGDYDRLGVNDEYDILGVKKLRLVIRGRNIAVITEAA
ANVRYQMSDISPVDIIEKRMLESMEKSKI"
2520..3449
/gene="CPE1005"
2520..3449
/gene="CPE1005"
/note="309 aa, similar to C-terminal of pir:575961
magnesium/cobalt transport protein sl10507 from
Synecocystis sp. (strain PCC 6803) (387 aa); 22.4%
identity in 263 aa overlap. 2 putative transmembrane
regions were found by PSORT."
/codon_start=1
```

aa); 62.9% identity in 116 aa overlap
CPEB1009"
/codon_start=1
/transl_table=11
/product="methylglucosyl synthase"
/protein_id="BAB80715.1"
/db_xref="GI:18144670"
/translation="MKIALIDKKKEEMIELANDFEDLSKHILVATGTTGLKIMON
TSLERKRCSGPLGDDDEIGAMVAVNHVDVMVIFLDRPLTAQCHEPDISALLLSDYVK
VPLATNESAKLIMADI"
7278. 7613
/gene="CPEB1010"
7278. 7613
/gene="CPEB1010"
/note="111 aa, no significant homology. Putative
N-terminal signal sequence and 2 putative transmembrane
regions were found by PSORT."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB80716.1"
/db_xref="GI:18144671"
/translation="MRGINIQMSNMLIVSTIILITKYLLENLILCKENKIGLTK
KIKKVIIFLITVSIIVIIIGITFPEYKRFILNVAISQIRLIVAFQTLFGIILFLF
INRKKYKVK"
7739. 7999
/gene="CPEB1011"

gene

Query Match 8.9%; Score 45.6; DB 1; Length 298050;
Best Local Similarity 51.5%; Pred. No. 1.4;
Matches 105; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 98 AATACGTTTCTCCAGTTGTATTAAGACCCGCCACCGATTAAGCTCATGCAAGA 157
Db 208707 AAGTTTTTTTATTAATTTTCTTAATAATATTCATATAAAATCAAGATTATTTCAATGTT 208766
QY 158 AGAAAATGTCATACATCTCTCTAGTCTCATATTATTTTCATTAGATAGCCGGTTTTT 217
Db 208767 AATAGTATTAAGATATCTTTGACATATATATTTATTTATATATATATTTAAT 208826
QY 218 ACTACACTCAATTAAGATGAACAGAAATGAATGGTTAGTGCTGTTTAAAGAAAGAT 277
Db 208827 TCTTATATATGATAGTAGGATGGTGAAGATGAGCAATATATACCTATTTTAAAGAAAAA 208886
QY 278 AATAAGACTACTCATCATTTGA 301
Db 208887 AATTAAACTAAGCTCTCAAGGA 208910

RESULT 3
AX083744 1141 bp DNA linear PAT 28-FEB-2001
LOCUS
DEFINITION Sequence 22 from Patent WO0111061.
ACCESSION AX083744
VERSION AX083744.1 GI:13185472
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
Kunst, L. and Clemens, S.
Regulation of embryonic transcription in plants
Patent: WO 011061-A 22 15-FEB-2001;
UNIVERSITY OF BRITISH COLUMBIA (CA)
Location/Qualifiers
1. 1141
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
1. 1141
/note="consensus sequence of A.t., L.a., and B.n. FAEI
promoters"

FEATURES
source
promoter

ORIGIN

Query Match	8.9%;	Score 45.4;	DB 6;	Length 1141;
Best Local Similarity	9.9%;	Pred. No. 4.4;		
Matches 46;	Conservative 197;	Mismatches 221;	Indels 0;	Gaps 0;

[illegible]

	AC094376	238301 bp	DNA linear	HTG_13-NOV-2002
RESULT 4	AC094376/c	Rattus norvegicus clone CH230-3Pl,	** SEQUENCING IN PROGRESS **	
LOCUS	DEFINITION	2 unordered pieces.		
ACCSSION	VERSION	AC094376		
KEYWORDS	SOURCE	AC094376..11 GI:24942492 HMG; HTGS_PHASEI; HTGS_DRAFT; HTGS_ENRICHED. Rattus norvegicus (Norway rat)		
ORGANISM		Rattus norvegicus		

REFERENCE
AUTHORS

1 (bases 1 to 238301)

Muzny, D., Marie, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amlin, A., Anguiano, D., Ayala-Bechli, V., Aoyagi, A., Ayodeji, M., Bacce, E., Bader, H., Bandaranaike, D., Bandaranaike, D., Barber, M., Barnescead, M., Benahmed, F., Biewalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Centea, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carrillo, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Dublin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.G., Gaisli, A., Ganta, R., Garcia, A., Garner, T., Gatzar, M., Georgeogis, E., Geer, K., Gill, R., Grady, M., Guerra, T., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulky, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kwis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, Y., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshuhwa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Nangum, A.,
 Nangum, B., Nagua, P., Martin, K., Martin, R., Matlinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Mlosoavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
 Narkavicius, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwokediemenoh, O., Okunonu, G., Olanugban, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfennkoch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puzoz, M., Quitor, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojase, A., Rose, M., Rose, R., Ruiz, S., J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaiz, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J.,
 Steidle, M., Strong, R., Sutton, A., Swack, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, M., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, U.,
 Wang, Q., Wang, S., Warren, J., Warren, L., Wei, X., White, F.,
 Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H. O.,
 Weinstock, G., and Gibbs, R. A.

Direct Submission
 Unpublished
 2 (bases 1 to 238301)
 Morley, K.C.
 Direct Submission
 Submitted (15-SEP-2001) Human Genome Sequencing Center. Department

JOURNAL TITLE	Unpublished
REFERENCE	2. (bases 1 to 238301)
AUTHORS	Worley, K.C.
JOURNAL TITLE	Direct Submission
REFERENCE	Submitted (15-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
AUTHORS	3. (bases 1 to 238301)
REFERENCE	Rat Genome Sequencing Consortium.
JOURNAL TITLE	Direct Submission
REFERENCE	Submitted (13-NOV-2002) Human Genome Sequencing Center, Department

COMMENT

On Nov 13, 2002 this sequence version replaced g1:23265811. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAWG
Center clone name: CH230-3p1
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229728 bases at least Q40
Consensus quality: 232177 bases at least Q30
Consensus quality: 233540 bases at least Q20
Estimated insert size: 237791; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Gendbank\_draft\_data.html).

```

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 237092: contig of 237092 bp in length
 * 237093 237192: gap of unknown length
 * 237193 238301: contig of 1109 bp in length.
 Location/Qualifiers
 1. 238301
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-3P1"

ORIGIN

Query Match 8.9%; Score 45.4; DB 2; Length 238301;
 Best Local Similarity 53.0%; Pred. No. 1.6; Mismatches 0; Gaps 0;
 Matches 97; Conservative 0; Indels 0;

QY 173 ATCTCTTAGTCTCATTATTATTTTATTAGATAGCCGGTTTCTACTCAACTCAATA 232
 |||||
 Db 152787 ATCTCTTTGTGTAATTAATGTAATAAATTTCCCTTGATTATACATGATTTTATGT 152728

QY 233 AGATGACGAGATGATGGGTTTGACGCTGTTTATAAGAGATATTAAGATCTATTC 292
 |||||
 Db 152727 GAGTCGACATTAATAATGTTAGTATGCTCCCAAAAAAATATATAATAAAT 152668

QY 293 ATCATTTAGGCAATAGGAGGAGAGATTTCAGCAACGCTGCTTCAAGTGGAAA 352
 |||||
 Db 152667 GGAATATGAGCAATCAAGAGAGACATGCAACACCTGACATGAAATATCAAA 152608

QY 353 CAA 355
 |||||
 Db 152607 GAA 152605

RESULT 5
 AC026485/c 69335 bp DNA linear HTG 22-MAR-2000
 LOCUS AC026485
 DEFINITION Homo sapiens clone RP11-24F22, LOW-PASS SEQUENCE SAMPLING.
 AC026485
 VERSION AC026485.1 GI:7280315
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 69335)
 Birren, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N., Birtenshaw, S., Baldwin, J., Barina, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campoliano, A., Cooke, P., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Coyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Liu, G., Locke, K., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneses, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

TITLE
JOURNAL
COMMENT

Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vasiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submission@genome.wi.mit.edu

Project Information

Center project name: L446
 Center clone name: 24_F_22

* NOTE: This record contains 88 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 709: contig of 709 bp in length
 * 1 809: gap of 100 bp
 * 810 1511: contig of 702 bp in length
 * 1512 1611: gap of 100 bp in length
 * 1612 2309: contig of 638 bp in length
 * 2310 2409: gap of 100 bp
 * 2410 3108: contig of 639 bp in length
 * 3109 3208: gap of 100 bp
 * 3209 3893: contig of 685 bp in length
 * 3894 3993: gap of 100 bp
 * 3994 4686: contig of 693 bp in length
 * 4687 4787: gap of 100 bp
 * 4787 5482: contig of 636 bp in length
 * 5483 5582: gap of 100 bp
 * 5583 6292: contig of 710 bp in length
 * 6293 6393: gap of 100 bp in length
 * 6393 7092: contig of 700 bp in length
 * 7093 7192: gap of 100 bp
 * 7193 7865: contig of 673 bp in length
 * 7866 7965: gap of 100 bp
 * 7966 8657: contig of 692 bp in length
 * 8658 8757: gap of 100 bp
 * 8758 9444: contig of 687 bp in length
 * 9445 9544: gap of 100 bp
 * 9545 10237: contig of 693 bp in length
 * 10238 10337: gap of 100 bp
 * 10338 11014: contig of 677 bp in length
 * 11015 11114: gap of 100 bp
 * 11115 11178: contig of 664 bp in length
 * 11779 11878: gap of 100 bp
 * 11879 12555: contig of 677 bp in length
 * 12556 12655: gap of 100 bp
 * 12656 13360: contig of 705 bp in length
 * 13361 13460: gap of 100 bp
 * 13461 14163: contig of 703 bp in length
 * 14164 14934: contig of 671 bp in length
 * 14935 15034: gap of 100 bp
 * 15035 15720: contig of 686 bp in length
 * 15721 15820: gap of 100 bp
 * 15821 16514: contig of 694 bp in length
 * 16515 17304: gap of 100 bp
 * 17304: contig of 690 bp in length

	*	45867	46530:	contig of 664 bp in length																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													</
--	---	-------	--------	----------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

TITLE
JOURNAL
COMMENT

Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneu, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Young, G., Zaitoun, J., Zemek, U., Zimmer, A. and Zody, M.

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L27588
Center clone name: 2362_F_20

Submitted (17-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

* NOTE: This record contains 64 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 688: contig of 688 bp in length
* 689 788: gap of 100 bp
* 789 1509: contig of 721 bp in length
* 1510 1609: gap of 100 bp
* 1610 2354: contig of 745 bp in length
* 2355 2454: gap of 100 bp
* 2455 3158: contig of 704 bp in length
* 3159 3258: gap of 100 bp
* 3259 3989: contig of 731 bp in length
* 3990 4089: gap of 100 bp
* 4090 4825: contig of 736 bp in length
* 4826 4925: gap of 100 bp
* 4926 5657: contig of 732 bp in length
* 5658 5757: gap of 100 bp
* 5758 6489: contig of 732 bp in length
* 6490 6589: gap of 100 bp
* 6590 7298: contig of 709 bp in length
* 7299 7398: gap of 100 bp
* 7399 8132: contig of 734 bp in length
* 8133 8232: gap of 100 bp
* 8233 8962: contig of 730 bp in length
* 8963 9062: gap of 100 bp
* 9063 9790: contig of 728 bp in length
* 9791 9890: gap of 100 bp
* 9891 10625: contig of 735 bp in length
* 10626 10725: gap of 100 bp
* 10726 11457: contig of 731 bp in length
* 11457 11557: gap of 100 bp
* 11557 12283: contig of 727 bp in length
* 12284 12383: gap of 100 bp
* 12384 13130: contig of 747 bp in length
* 13131 13230: gap of 100 bp
* 13231 13983: contig of 753 bp in length
* 13984 14083: gap of 100 bp
* 14084 14832: contig of 749 bp in length
* 14833 14932: gap of 100 bp

* 14933 15661: contig of 729 bp in length
* 15662 15761: gap of 100 bp
* 15762 16482: contig of 721 bp in length
* 16483 16582: gap of 100 bp
* 16583 17298: contig of 716 bp in length
* 17299 17398: gap of 100 bp
* 17399 18144: contig of 746 bp in length
* 18145 18244: gap of 100 bp
* 18245 18799: contig of 735 bp in length
* 18800 19079: gap of 100 bp
* 19080 19797: contig of 718 bp in length
* 19798 19898: gap of 100 bp
* 19898 20621: contig of 724 bp in length
* 20622 20721: gap of 100 bp
* 20722 21471: contig of 749 bp in length
* 21471 21570: gap of 100 bp
* 21571 22316: contig of 746 bp in length
* 22317 22416: gap of 100 bp
* 22417 23125: contig of 709 bp in length
* 23126 23325: gap of 100 bp
* 23226 23968: contig of 743 bp in length
* 23969 24068: gap of 100 bp
* 24069 24794: contig of 726 bp in length
* 24795 24894: gap of 100 bp
* 24895 25621: contig of 727 bp in length
* 25622 25721: gap of 100 bp
* 25722 26453: contig of 732 bp in length
* 26454 26553: gap of 100 bp
* 26554 27300: contig of 747 bp in length
* 27301 27400: gap of 100 bp
* 27401 28125: contig of 725 bp in length
* 28126 28825: gap of 100 bp
* 28826 29074: contig of 749 bp in length
* 28975 29074: gap of 100 bp
* 29075 29816: contig of 742 bp in length
* 29817 29916: gap of 100 bp
* 29917 30663: contig of 747 bp in length
* 30664 30763: gap of 100 bp
* 30764 31497: contig of 734 bp in length
* 31498 31597: gap of 100 bp
* 31598 32331: contig of 734 bp in length
* 32332 32431: gap of 100 bp
* 32432 33159: contig of 728 bp in length
* 33160 33259: gap of 100 bp
* 33260 33988: contig of 729 bp in length
* 33989 34088: gap of 100 bp
* 34089 34835: contig of 747 bp in length
* 34836 34935: gap of 100 bp
* 34936 35671: contig of 736 bp in length
* 35672 35771: gap of 100 bp
* 35772 36505: contig of 734 bp in length
* 36506 36605: gap of 100 bp
* 36606 37286: contig of 681 bp in length
* 37287 37386: gap of 100 bp
* 37387 38115: contig of 729 bp in length
* 38116 38215: gap of 100 bp
* 38216 38944: contig of 729 bp in length
* 38945 39044: gap of 100 bp
* 39045 39765: contig of 721 bp in length
* 39766 39865: gap of 100 bp
* 39866 40599: contig of 734 bp in length
* 40600 40699: gap of 100 bp
* 40700 41441: contig of 742 bp in length
* 41442 41541: gap of 100 bp
* 41542 42257: contig of 616 bp in length
* 42257 42517: gap of 100 bp
* 42518 42989: contig of 732 bp in length
* 42990 43089: gap of 100 bp
* 43090 43805: contig of 716 bp in length
* 43806 43905: gap of 100 bp
* 43906 44654: contig of 749 bp in length
* 44655 44754: gap of 100 bp
* 44755 45498: contig of 744 bp in length

45499: gap of 100 bp
45599: contig of 737 bp in length
46335: gap of 100 bp
46336: gap of 100 bp
46436: contig of 714 bp in length
47149: gap of 100 bp
47250: contig of 731 bp in length
47250: contig of 731 bp in length
47980: gap of 100 bp
47981: contig of 735 bp in length
48080: gap of 100 bp
48820: gap of 100 bp
48919: gap of 100 bp
48920: gap of 100 bp
49646: contig of 727 bp in length
49746: gap of 100 bp
50474: contig of 727 bp in length
50573: gap of 100 bp
51268: contig of 695 bp in length
51269: gap of 100 bp
51369: contig of 726 bp in length
52094: gap of 100 bp
52195: contig of 721 bp in length.
52195: Location/Qualifiers
1. 52915
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="CTD-2362F20"
/clone_1id="CITDI Human BAC"

ORIGIN

Query Match 8.7%; Score 44.4; DB 2; Length 52915;
Best Local Similarity 50.0%; Pred. No. 3.6; Indels 5; Gaps 1;

Matches 141; Conservative 0; Mismatches 136; Indels 5; Gaps 1;

QY 19 ATGCTTATTTATGTCAGTAAATGATTTCTCCCAATTAAGTGAAGAGAGAT 78
DB 39896 ATCTCGTCGACGTGAGATGAATGACATTTTAAAGACAAAGTAAAGTCCAGAGAGAT 39955
QY 79 TGTGTTTATCTGTGGTAAATACGTTTCTCCAGTTGATTAAGACCTCCACCACT 138
DB 39956 AAATTCAGTCTATCTCAATTAATTTCTATCAAGAACTGGGAAACTTCCA----- 40010
QY 139 ATAAAGTCTATGTCAGCAAGAAATGTCATATCATTCTTGTCTCATTATTATTTTC 198
DB 40011 AATTAAGTCTATGTCAGCAAGAAATGTCATATCATTCTTGTCTCATTATTATTTTC 198
QY 199 ATTAGATGCGCGGTTTCTTCTCACTCAATTAAGTGAAGAGATGAGTGTAGT 258
DB 40071 AAAAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 40130
QY 259 ACTGTTTATTAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 300
DB 40131 ACTCTCAAACTACAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 40172

RESULT 7

PFMAL13PB/c 83110 bp DNA linear HTG 11-AUG-1999
LOCUS Plasmodium falciparum 3D7 chromosome 13, *** SEQUENCING IN PROGRESS
DEFINITION *** 9 unordered pieces.

ACCESSION AL109814.1 GI:5731891
VERSION AL109814.1
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 Bowman, S., Churcher, C., Harris, B., Harris, D., Lawson, D., Quail, M.
and Barrett, B.
Direct Submission
Submitted (11-AUG-1999) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK

COMMENT

For more information about this sequence or the Malaria Project,

see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
sequence is unfinished and does not necessarily represent the
correct sequence. Work on the sequence is in progress and the
release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
Order of segments is not known, 800 n's separate segments.
NOTE: This is a 'working draft' sequence. It currently
consists of 9 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 12609: contig of 12609 bp in length
* 12610 13409: gap of 800 bp
* 13410 15875: contig of 2466 bp in length
* 15876 16675: gap of 800 bp
* 16676 21699: contig of 5024 bp in length
* 21700 22499: gap of 800 bp
* 22500 25196: contig of 2697 bp in length
* 25197 25996: gap of 800 bp
* 25997 28165: contig of 2168 bp in length
* 28166 28965: gap of 800 bp
* 28966 37197: contig of 8233 bp in length
* 37198 37997: gap of 800 bp
* 37998 41776: contig of 3779 bp in length
* 41777 42577: gap of 800 bp
* 42577 47712: contig of 5135 bp in length
* 47712 48511: gap of 800 bp
* 48512 83110: contig of 34599 bp in length.

FEATURES

source
1. 83110
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/bctain="3D7"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="13"

ORIGIN

Query Match 8.6%; Score 43.8; DB 2; Length 83110;
Best Local Similarity 46.3%; Pred. No. 4.6;
Matches 144; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 TGGATCCTTCATCTCATGCTTATATGTCAGATGATGATTTCTCAACCAATTAAGT 63
DB 598 TGTGTTTCTTCAACAAGTATATTAATTAATTAATTAATTAATTAATTAATTAATTA 539
QY 64 GACAAAGCAGAGATTTGTTTATCTGTGGTAAATACGTTTCTCCAGTTGATTA 123
DB 538 AACAAATTAATTTTATTTTATTAAGATGTAAGATTAATTAATTAATTAATTAATTA 479
QY 124 GACCTCCACAGATTAAGTCTATGCAACAAAGAAATGTCATTAATCTCTTAGT 183
DB 478 AATCTATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 419
QY 184 CTCATATATTTTCTATGATAGTCCGTTTCTTCTACCAATTAATTAAGATGAGAC 243
DB 418 TTTTATTTTATCTTATTTTCTGATTTTATTAATTAATTAATTAATTAATTAATTA 359
QY 244 ATGAATGAGTATGATGATGATTAATTAAGAGATTAATTAAGATTAATTAATTAATTA 303
DB 358 AAAAATGACAAAGAAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 259
QY 304 CAATTAAGGAG 314
DB 298 ATAAAGGCG 288

RESULT 8

PFMAL13_13
WPCOMMENT
Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509
Fragment Name Begin End
PFMAL13_00 1 110000
PFMAL13_01 100001 210000
PFMAL13_02 200001 310000
PFMAL13_03 300001 410000
PFMAL13_04 400001 510000
PFMAL13_05 500001 610000
PFMAL13_06 600001 710000
PFMAL13_07 700001 810000
PFMAL13_08 800001 910000
PFMAL13_09 900001 1010000
PFMAL13_10 1000001 1110000
PFMAL13_11 1100001 1210000
PFMAL13_12 1200001 1310000
PFMAL13_13 1300001 1410000
PFMAL13_14 1400001 1510000
PFMAL13_15 1500001 1610000
PFMAL13_16 1600001 1710000
PFMAL13_17 1700001 1810000
PFMAL13_18 1800001 1910000
PFMAL13_19 1900001 2010000
PFMAL13_20 2000001 2110000
PFMAL13_21 2100001 2210000
PFMAL13_22 2200001 2310000
PFMAL13_23 2300001 2410000
PFMAL13_24 2400001 2510000
PFMAL13_25 2500001 2610000
PFMAL13_26 2600001 2710000
PFMAL13_27 2700001 2732359
Continuation (14 of 28) of PFMAL13 from base 1300001 (AL844509 Plasmidium falciiparum 3D7)

Query Match 8.6%; Score 43.8; DB 2; Length 110000;
Best Local Similarity 46.3%; Pred. No. 4.3;
Matches 144; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 TGGATCCTTCATCTCATGCTTATTATGTGAGTGAATAGATTCTGACCAATTGAAAG 63
DB 103932 TGTGTTTTCTTCAACAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 103991
QY 64 GACAAAGCAGAGATTGTTGTTTATCTGTTGGTAAATACGTTTCCAGTTGATAAA 123
DB 103992 AACAAATTAATTTTTTTTTTTTTTTTATTAAGAGTAAAAAAATATATGTTTACATATA 104051
QY 124 GACCTCCCAACGATTAAGTCCATGCAACAAAGAAAAATGCAATACATCTCTTACT 183
DB 104052 AATCTATATATGATTAATAAATCTATGACACCCCTAAATTTAGATTAATTTTTTTT 104111
QY 184 CTCATTATATTCTTCATAGATAGCCGTTTTTTTACTACACTCACTCAATAAGATGACAGA 243
DB 104112 TTTTCTTTTATCTTAATTTTTTTCGTCATTTTTTATTATTAACCTAAAAA 104171
QY 244 ATGAATGGGTTAGTGACTGTTTATAAAGAGATTAATAAGATCATCATTTGAGG 303
DB 104172 AAAAAATGACAAAGAAATGAATAATTAATTAATTAATTAATTAATTAATTAATTA 104231
QY 304 CAATTAAGGAG 314
DB 104232 ATAAAGGGCG 104242

RESULT 9
PFMAL13_14
WPCOMMENT
Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509
Fragment Name Begin End
PFMAL13_00 1 110000
PFMAL13_01 100001 210000
PFMAL13_02 200001 310000
PFMAL13_03 300001 410000
PFMAL13_04 400001 510000

PFMAL13_05 500001 610000
PFMAL13_06 600001 710000
PFMAL13_07 700001 810000
PFMAL13_08 800001 910000
PFMAL13_09 900001 1010000
PFMAL13_10 1000001 1110000
PFMAL13_11 1100001 1210000
PFMAL13_12 1200001 1310000
PFMAL13_13 1300001 1410000
PFMAL13_14 1400001 1510000
PFMAL13_15 1500001 1610000
PFMAL13_16 1600001 1710000
PFMAL13_17 1700001 1810000
PFMAL13_18 1800001 1910000
PFMAL13_19 1900001 2010000
PFMAL13_20 2000001 2110000
PFMAL13_21 2100001 2210000
PFMAL13_22 2200001 2310000
PFMAL13_23 2300001 2410000
PFMAL13_24 2400001 2510000
PFMAL13_25 2500001 2610000
PFMAL13_26 2600001 2710000
PFMAL13_27 2700001 2732359
Continuation (15 of 28) of PFMAL13 from base 1400001 (AL844509 Plasmidium falciiparum 3D7)

Query Match 8.6%; Score 43.8; DB 2; Length 110000;
Best Local Similarity 46.3%; Pred. No. 4.3;
Matches 144; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 TGGATCCTTCATCTCATGCTTATTATGTGAGTGAATAGATTCTGACCAATTGAAAG 63
DB 3932 TGTGTTTTCTTCAACAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3991
QY 64 GACAAAGCAGAGATTGTTGTTTATCTGTTGGTAAATACGTTTCCAGTTGATAAA 123
DB 3992 AACAAATTAATTTTTTTTTTTTTTTTATTAAGAGTAAAAAAATATATGTTTACATATA 4051
QY 124 GACCTCCCAACGATTAAGTCCATGCAACAAAGAAAAATGCAATACATCTCTTACT 183
DB 4052 AATCTATATATGATTAATAAATCTATGACACCCCTAAATTTAGATTAATTTTTTTT 4111
QY 184 CTCATTATATTCTTCATAGATAGCCGTTTTTTTACTACACTCACTCAATAAGATGACAGA 243
DB 4112 TTTTCTTTTATCTTAATTTTTTTCGTCATTTTTTATTATTAACCTAAAAA 4171
QY 244 ATGAATGGGTTAGTGACTGTTTATAAAGAGATTAATAAGATCATCATTTGAGG 303
DB 4172 AAAAAATGCAAAAGAAATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4231
QY 304 CAATTAAGGAG 314
DB 4232 ATAAAGGGCG 4242

RESULT 10
AL136371
LOCUS AL136371 147784 bp DNA linear PRI 30-NOV-2000
DEFINITION Human DNA sequence from clone Rpl1-110P20 on chromosome 1q31.1
ACCESSION AL136371
VERSION AL136371.8 GI:9187156
KEYWORDS HTG; Cpg Island.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 147784)
AUTHORS Chapman, J.
TITLES Direct Submission
JOURNAL Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT On Jul 14, 2000 this sequence version replaced gi:8894185.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr1>

RP11-110P20 is from the library RP11-11.1 constructed at the

Roswell Park Cancer Institute by the group of Pieter de Jong. For

further details see <http://bacpac.med.buffalo.edu/>

VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-110P20 The true

left end of clone RP11-487023 is at 59783 in this sequence.

Location/Qualifiers

```
source
1..147784
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q31.1"
/clone="RP11-110P20"
/clone_1lb="RP11-11.1"
8..539
/note="match: GSS: Em:AQ321011"
repeat_region
181..447
/note="L1MA7 repeat: matches 6014..6255 of consensus"
misc_feature
202..675
/note="match: GSS: Em:AQ406263"
repeat_region
752..956
/note="MER58A repeat: matches 1..213 of consensus"
repeat_region
1045..1214
/note="MER58A repeat: matches 1..175 of consensus"
repeat_region
1887..1990
/note="L2 repeat: matches 2562..2672 of consensus"
repeat_region
3724..3763
/note="20 copies 2 mer ac 77% conserved"
repeat_region
3917..4210
/note="AluY repeat: matches 1..294 of consensus"
5674..6134
/note="MT1H repeat: matches 68..512 of consensus"
repeat_region
7252..7537
/note="AluX repeat: matches 10..296 of consensus"
complement(7552..7954)
/note="match: GSS: Em:AQ018563"
8079..8200
/note="match: GSS: Em:A2040820"
8206..8595
/note="match: GSS: Em:A2040820"
repeat_region
9363..9625
/note="L1MC4 repeat: matches 7306..7586 of consensus"
9652..10604
/note="L1M1 repeat: matches 1390..390 of consensus"
10596..10861
/note="L1M1 repeat: matches 4665..4914 of consensus"
10849..13488
/note="L1PA13 repeat: matches 3488..6156 of consensus"
13497..13917
/note="L1M1 repeat: matches 4906..5255 of consensus"
13918..14162
repeat_region
```

```
/note="AluSg repeat: matches 3..247 of consensus"
14163..14306
/note="L1M1 repeat: matches 5255..5403 of consensus"
14303..15870
/note="L1PA7 repeat: matches 4578..6141 of consensus"
17029..17153
/note="MIR repeat: matches 48..183 of consensus"
complement(17194..17431)
/note="match: GSS: Em:AQ627818"
complement(17203..17626)
/note="match: GSS: Em:AQ382224"
17205..17574
/note="match: GSS: Em:AQ344420"
17239..17623
/note="MER4B repeat: matches 206..574 of consensus"
17641..17749
/note="AluSg/x repeat: matches 178..286 of consensus"
17751..18087
/note="MER4D repeat: matches 236..584 of consensus"
18188..18482
/note="AluY repeat: matches 1..296 of consensus"
18630..18980
/note="MER4B repeat: matches 1..236 of consensus"
19636..19847
/note="MER20 repeat: matches 2..217 of consensus"
21147..22145
/note="PTR5 repeat: matches 744..2021 of consensus"
21337..21759
/note="CpG island"
/evidence="not_experimental"
22150..22247
/note="L1R30 repeat: matches 544..642 of consensus"
22257..22471
/note="PTR5 repeat: matches 177..392 of consensus"
22741..22869
/note="L1P1 repeat: matches 5681..5818 of consensus"
22841..22895
/note="L1PA3 repeat: matches 6092..6146 of consensus"
23141..23286
/note="AluSg/x repeat: matches 1..115 of consensus"
23402..23802
/note="match: GSS: Em:AQ797421"
complement(23554..24004)
/note="match: GSS: Em:AQ007707"
23816..23939
/note="62 copies 2 mer aa 58% conserved"
24049..24742
/note="L1MC4 repeat: matches 6555..7295 of consensus"
24800..25176
/note="MER92A repeat: matches 1..359 of consensus"
24919..25385
/note="match: GSS: Em:AQ675810"
25096..25485
/note="match: GSS: Em:B38318"
25198..25276
/note="TIGGR2 repeat: matches 2628..2708 of consensus"
25279..25332
/note="L1PA16 repeat: matches 6103..6157 of consensus"
25521..30992
/note="L1PA5 repeat: matches 688..6157 of consensus"
31192..31791
/note="L1PA13 repeat: matches 651..50 of consensus"
31801..32118
/note="TIGGR2 repeat: matches 2338..2658 of consensus"
32148..32216
/note="TIGGR2 repeat: matches 1..69 of consensus"
32235..32572
/note="AluY repeat: matches 1..307 of consensus"
33626..33978
/note="MER47B repeat: matches 11..418 of consensus"
34298..34625
/note="MER2 repeat: matches 1..341 of consensus"
34950..34985
repeat_region
```

```
misc_feature /note="18 copies 2 mer ct 80% conserved"
complement(35544..35956)
/note="match: GSS: Em:AQ184339"
repeat_region 36136..36231
/note="6 copies 16 mer 72% conserved"
repeat_region 36145..36230
/note="43 copies 2 mer ca 76% conserved"
repeat_region 37001..37267
/note="11usC repeat: matches 32..298 of consensus"
repeat_region 37814..37912
/note="12 repeat: matches 2608..2710 of consensus"
repeat_region 38718..38839
/note="11 repeat: matches 4706..4841 of consensus"
repeat_region 39077..39381
/note="11usP repeat: matches 1..313 of consensus"
repeat_region 39552..39750
/note="11M/C/D repeat: matches 5281..5479 of consensus"
repeat_region 40657..40784
/note="8 copies 16 mer 68% conserved"
repeat_region 40698..40785
/note="44 copies 2 mer ag 64% conserved"
misc_feature complement(40973..41511)
/note="match: GSS: Em:B55962"
misc_feature complement(41075..41505)
/note="match: GSS: Em:AQ217602"
misc_feature complement(41254..41503)
/note="match: GSS: Em:AQ076892"
complement(41291..41511)
/note="match: GSS: Em:AQ09585"
repeat_region 41765..41965
/note="match: GSS: Em:AQ09585"
/note="11M/A7 repeat: matches 6075..6283 of consensus"
misc_feature 42059..42402
/note="match: GSS: Em:AQ215516"
misc_feature 42059..42255
/note="match: GSS: Em:B46341"
misc_feature 42089..42255
/note="match: GSS: Em:AQ036131"
```

```
Query Match 8 6%; Score 43.8; DB 9; Length 147784;
Best Local Similarity 53.8%; Pred. No. 4.1; Indels 2; Gaps 2;
Matches 133; Conservative 0; Mismatches 112; Indels 2; Gaps 2;

Qy 39 AATAGATTCACCAATTAAGATGACAAAGCAGATTTGTTTATCTGTGGGTA 98
Db 35138 ACTTCACCTTTACCAAGTGAATTAATAAATTTGTAAGAAATATTTCTTTTA 35197

Qy 99 AATACGTTTTCACAGTTGTAAT-AAAGCCCTCCACAGTATAAAGCTTATGCAACA 157
Db 35198 AAGTTAGTTTCTACTTATATGAGAGTCATTAACAATTTTAAAGTTTATGTTATA 35257

Qy 158 AGAAATGTCATACATTCCTTAGTCTATTTAT-TTTCTTAGATAGCCGGTTTTT 216
Db 35258 AGGTCTTGAAAGAAATGCTTGGCCATATATTTGTTTCATTAATCTGTTAG 35317

Qy 217 TACTCAACTCAATTAAGATGAACAGATGATGGTTAGTGTGTTATATAAGAGAG 276
Db 35318 TCTTATATCTCAGTGTGATTAAGACCTGAAGGATTAATTTGATGATGGAAAA 35377

Qy 277 TAATATA 283
Db 35378 TCTTAGA 35384
```

RESULT 11
PFMALIP2_2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMALIP2 Accession AL031745

```
Fragment Name Begin End
PFMALIP2_0 1 110000
PFMALIP2_1 100001 210000
PFMALIP2_2 200001 310000
PFMALIP2_3 300001 384550
Continuation (3 of 4) of PFMALIP2 from base 200001 (AL031745 Plasmidium falciptarum DNA B
```

```
Query Match 8 5%; Score 43.6; DB 3; Length 110000;
Best Local Similarity 51.0%; Pred. No. 4.8; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 17 TCATGCTTATATATGAGTGAATGATGATTCACCAATTAAGATGACAAAGCAGAGA 76
Db 47331 TTATCCCTTTATGGGGTTAAAAAATATACATTAATAAAGAAAAAGAAA 47332

Qy 77 TTTGTTTTATCTGTTGGGTAATACGTTTTCAGTTGTATTAAGACCTCCACCA 136
Db 47331 AACCGATCTATTTTATAGTATAAAAAATATATTAATTAATAATATTTATPA 47272

Qy 137 GTATTAAGTCCATGACCAAAAGAAATGCAATCATTCCTTGTCTCATATATT 196
Db 47271 GTAATCTTATATTAATAATATATATTCATTCATTCATGACATATTTTCTA 47212

Qy 197 TCATTAGATAGCCGGTTTTTA 218
Db 47211 GAACAATATAGCTATATTATTA 47190
```

RESULT 12
CR450691 149266 bp DNA linear HTG 24-MAY-2004
LOCUS CR450691
DEFINITION Danio rerio clone CH211-77B10, WORKING DRAFT SEQUENCE, 6 unordered
pieces
ACCESSION CR450691.1 GI:47604349
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
KEYWORDS Danio rerio (zebrafish)
SOURCE Danio rerio
ORGANISM Danio rerio

REFERENCE
AUTHORS Burton, J.
TITLE Submitted (21-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
JOURNAL Genome Center
COMMENT Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
Project Information
Center project name: zc77B10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: dye-terminator; 100% of reads
Consensus quality: 146835 bases at least Q40
Consensus quality: 147357 bases at least Q30
Consensus quality: 147782 bases at least Q20
Insert size: 148766; sum-of-contigs
Insert size: 158637; 6.4% error; agarose-fp
Quality coverage: 7.19x in Q20 bases; sum-of-contigs Quality
coverage: 6.80x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
* 1 8085: contig of 8085 bp in length
* 8086 8185: gap of 100 bp
* 8186 48371: contig of 40186 bp in length
* 48372 48471: gap of 100 bp
* 48472 52239: contig of 3768 bp in length
* 52240 52339: gap of 100 bp
```

* 52340 113800: contig of 61461 bp in length
 * 113801 113900: gap of 100 bp
 * 113901 140425: contig of 26525 bp in length
 * 140426 140525: gap of 100 bp
 * 140526 149266: contig of 8741 bp in length.
 Location/Qualifiers

FEATURES

source 1.149266
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-77E10"
 /clone_11b="CHORI-211"
 1..8085
 /note="assembly_fragment:00038
 clone_end:SP6
 vector_side:left"
 misc_feature 8186..48371
 /note="assembly_fragment:00595
 fragment_chain:1"
 misc_feature 48472..52239
 /note="assembly_fragment:00016
 fragment_chain:1"
 misc_feature 52340..113800
 /note="assembly_fragment:01002
 fragment_chain:1"
 misc_feature 113901..140425
 /note="assembly_fragment:00257
 fragment_chain:1"
 misc_feature 140526..149266
 /note="assembly_fragment:00146
 clone_end:T7
 vector_side:right"

ORIGIN

Query Match 8.5%; Score 43.6; DB 2; Length 149266;
 Best Local Similarity 54.3%; Pred. No. 4.5;
 Matches 88; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
 QY 62 TGGACAAAGCAGATTTGTTTATCTGTTGGGTAATACGTTTCTCCAGTTGATA 121
 DB 126007 TTGAAAAGCAGCAATTTTGTGTTAAATTTTAAATTAATTAATTCACAAAGCTGATT 126066
 QY 122 AGACCTCCACACAGATTAATCTCATGCAAGAAATGTCATCATCTCTTA 181
 DB 126067 TGAATGTTTATAGTTATATCAAAACTTACTATACATGATATATATGACTTG 126126
 QY 182 GTCCTATTATTATTCATTAGATCCGGTTTTTACTACA 223
 DB 126127 ATTTTATTATTATTTTAAATAATTTTACATTTTTACA 126168

RESULT 13
 AL390999/c 164505 bp DNA 1linear HTG 10-JUL-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-151E7, 17 unordered pieces.
 DEFINITION AL390999
 ACCESSION AL390999.3 GI:10178785
 VERSION HTG; HTGS_PHASE1; HTGS_CANCELLED.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS McIay, K.
 TITLE Direct Submmission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Sep 16, 2000 this sequence version replaced gi:10040136.
 COMMENT ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk

FEATURES

source

Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA151E7
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 158161 bases at least Q40
 Consensus quality: 160871 bases at least Q30
 Consensus quality: 161986 bases at least Q20
 Insert size: 162905; sum-of-contigs
 Insert size: 168883; 4.3% error; agarose-fp
 Quality coverage: 4.44x in Q20 bases; sum-of-contigs Quality
 Coverage: 4.32x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 3194: contig of 3194 bp in length
 * 3195 3284: gap of 100 bp
 * 3285 7213: contig of 3919 bp in length
 * 7214 7313: gap of 100 bp
 * 7314 18523: contig of 11210 bp in length
 * 18524 18623: gap of 100 bp
 * 18624 25335: contig of 6712 bp in length
 * 25336 25435: gap of 100 bp
 * 25436 30801: contig of 5366 bp in length
 * 30802 30901: gap of 100 bp
 * 30902 39872: contig of 8971 bp in length
 * 39873 39972: gap of 100 bp
 * 39973 52124: contig of 12152 bp in length
 * 52125 52224: gap of 100 bp
 * 52225 57860: contig of 5636 bp in length
 * 57861 57960: gap of 100 bp
 * 57961 75786: contig of 17826 bp in length
 * 75787 75886: gap of 100 bp
 * 75887 100086: contig of 24700 bp in length
 * 100087 100186: gap of 100 bp
 * 100187 107095: contig of 6909 bp in length
 * 107096 107195: gap of 100 bp
 * 107196 125498: contig of 18303 bp in length
 * 125499 125598: gap of 100 bp
 * 125599 128838: contig of 3240 bp in length
 * 128839 128938: gap of 100 bp
 * 128939 147758: contig of 18820 bp in length
 * 147759 147858: gap of 100 bp
 * 147859 150588: contig of 2730 bp in length
 * 150589 150688: gap of 100 bp
 * 150689 162243: contig of 11555 bp in length
 * 162244 162343: gap of 100 bp
 * 162344 164505: contig of 2162 bp in length.
 Location/Qualifiers
 1..164505
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-151E7"
 /clone_11b="RPC1-11.1"
 1..3194
 /note="assembly_fragment:00973
 fragment_chain:1
 clone_end:SP6
 vector_side:left"
 misc_feature 3295..7213
 /note="assembly_fragment:00415
 fragment_chain:1"
 misc_feature 7314..18523


```

/note="assembly fragment:01529
fragment chain:1"
misc_feature
/note="assembly fragment:01161
fragment chain:1"
misc_feature
/note="assembly fragment:00848
fragment chain:1"
misc_feature
/note="assembly fragment:00318
fragment chain:2"
misc_feature
/note="assembly fragment:00647
fragment chain:2"
misc_feature
/note="assembly fragment:00648
fragment chain:3"
misc_feature
/note="assembly fragment:01018
fragment chain:3"
misc_feature
/note="assembly fragment:00762
fragment chain:4"
misc_feature
/note="assembly fragment:00263
fragment chain:4"
misc_feature
/note="assembly fragment:00161"
misc_feature
/note="assembly fragment:00164"
misc_feature
/note="assembly fragment:00664"
misc_feature
/note="assembly fragment:01387"
misc_feature
/note="assembly fragment:01428"
misc_feature
/note="assembly fragment:01491
clone end:T7
vector_side:right"
ORIGIN
Query Match
Best Local Similarity 8.5%; Score 43.6; DB 2; Length 164505;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 138 TATTAAGTCTATGCACAAAGAAATGTCATACATCTCTTCTGCTCATTTATTTT 197
DB 125236 TATTAAGTCTATGCACAAAGAAATGTCATACATCTCTTCTGCTCATTTATTTT 125177
QY 198 CATTAGATGCGCGTTTCTTACTACAACTCAATTAAGATGAACAGAAATGCGTTAGT 257
DB 125176 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 125117
QY 258 GACTGTTTATTAAGAAAGATTAATTAAGATTAATTAATTAATTAATTAATTAATTAAT 291
DB 125116 GACTTATTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 125083
RESULT 14
AC024501/c AC024501 164595 bp DNA linear HTG 04-MAY-2001
LOCUS Homo sapiens chromosome 1 clone RP11-151E7 map 1, WORKING DRAFT
DEFINITION
SEQUENCE 14 unordered pieces.
ACCESSION AC024501
VERSION AC024501.2 GI:9994152
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 164595)
AUTHORS Birren,B., Linton,L., Nuebaum,C. and Lander,B.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Homo sapiens chromosome 1, clone RP11-151E7
2 (bases 1 to 164595)
Birren,B., Linton,L., Nuebaum,C., Lander,B., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boudgallier,B., Brown,A., Burkett,G., Campiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karkates,A.,
Klein,J., Lander,B., Laroque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N., McCarthy,M.,
McEwan,P., McGuck,A., McKernan,K., McPheters,R., Meldrum,J.,
Menues,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivat,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J., Tittel,A.,
Travers,M., Trigilio,J., Vasiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 8, 2000 this sequence version replaced gi:7108297.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 151.E7
Center clone name: 151.E7
----- Summary Statistics
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156974 bases at least Q40
Consensus quality: 16134 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 163295; sum-of-contrigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1757 1756: contig of 1756 bp in length
1857 1856: gap of 100 bp
35722 35722: contig of 3366 bp in length
35723 35722: gap of 100 bp
35823 38369: contig of 2547 bp in length
38370 38469: gap of 100 bp
38470 42940: contig of 4471 bp in length
42941 43040: gap of 100 bp
43041 48601: contig of 5561 bp in length
48602 48701: gap of 100 bp
58127 58127: contig of 9426 bp in length
58227 58227: gap of 100 bp
58228 68052: contig of 9855 bp in length
68053 68152: gap of 100 bp

```


Oy 258 GACTGTTTATAAAGAGATTAATAAGATAT 291
||| | | | | | | | | | | | | | |
Db 98915 GACTTTATAGTAGAATAATGCTAAATCTAATT 98882

Search completed: November 21, 2005, 07:00:36
Job time : 2422.28 secs

1	41.8	12.4	500	1	AU088319	AU088319	AU088319
2	41.4	12.3	444	5	BM339785	BM139785	BM139785
3	41.4	12.3	504	5	BM289615	BM289615	BM289615
4	41.4	12.3	574	1	AV680413	AV680413	AV680413
5	41.4	12.3	720	5	BM135968	BM135968	BM135968
6	41.4	12.3	725	5	BM186283	BM186283	BM186283
7	41.4	12.3	730	5	BM140184	BM140184	BM140184
8	41.4	12.3	722	5	BM127139	BM127139	BM127139
9	41.4	12.3	759	1	AV877102	AV877102	AV877102
10	41.4	12.3	767	5	BM082303	BM082303	BM082303
11	41.4	12.3	774	5	BM141206	BM141206	BM141206
12	41	12.2	598	7	CE250818	CE250818	CE250818
13	40.6	12.1	594	8	AZ305625	AZ305625	AZ305625
14	40.6	12.1	732	9	AG596239	MUS musc	AG596239
15	40.6	12.1	781	9	AG470649	MUS musc	AG470649
16	40.4	12.0	554	5	BM009156	BM009156	BM009156
17	40.2	12.0	552	8	AZ431375	AZ431375	IM021611
18	40	11.9	879	8	BH137539	BH137539	BH137539
19	40	11.9	943	8	BH147340	BH147340	ENTNA837T
20	39.8	11.8	472	5	BM093545	BM093545	BH147340
21	39.8	11.8	507	5	BM234156	BM234156	ENTQW347T
22	39.8	11.8	688	5	BM133492	BM133492	BM093545
23	39.8	11.8	760	5	BM139286	BM139286	BM234156
24	39.2	11.7	1101	9	CNS00E06	AL069591	BM133492
						Drosophila	BM093296
							AL069591

C	25	38.8	11.5	611	9	CE141435	CE141435	t1gr-gss-
C	26	38.8	11.5	759	9	BM403626	BM403626	
C	27	38.8	11.5	890	8	BH135160	BH135160	ENTROB35TF
C	28	38.6	11.5	463	5	BM124142	BM124142	BM124142
C	29	38.6	11.5	630	5	BW254209	BW254209	BW254209
C	30	38.6	11.5	715	8	BZ393910	BZ393910	EINW668TR
C	31	38.6	11.5	932	8	BH146427	BH146427	ENTPH922TR
C	32	38.6	11.5	987	9	CNS01490	AL104456	Drosophyll1
C	33	38.4	11.4	430	2	BB784350	BB784350	BB784350
C	34	38.4	11.4	665	8	BH973910	BD452e01.	
C	35	38.4	11.4	715	8	BZ011773	BZ011773	ceh48d03.
C	36	38.4	11.4	848	9	CR114143	CR114143	Reverse s
C	37	38	11.3	550	7	CK383193	CK383193	1a117908.
C	38	37.8	11.2	331	8	AZ002551	RPCT-23-2	
C	39	37.8	11.2	478	2	BZ786550	EST451141	BZ786550
C	40	37.8	11.2	905	4	AF249279	ENTEG46TF	
C	41	37.8	11.2	928	9	CNS01761	AL107995	Drosophyll1
C	42	37.8	11.2	1200	9	CNS01651	AL106527	Drosophyll1
C	43	37.6	11.2	489	9	CLB81587	CLB81587	abf1sc07.
C	44	37.6	11.2	586	8	BZ948176	BZ948176	
C	45	37.6	11.2	747	6	CD825484	BN25.0650	

ALIGNMENTS

RESULT 1	
AY0088319	500 bp mRNA linear EST 27-JAN-2001
LOCUS	
DEFINITION	AY0088319 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
ACCESSION	clone XPfn6924, mRNA sequence.
	AY0088319

SOURCE

REFERENCE	AUTHORS	TITLE	JOURNAL
Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmodium. 1 (bases 1 to 500)	Watanabe, J., Sasaki, M., Suzuki, Y. and Sugano, S.	FULL-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum	Nucleic Acids Res. 29 (1), 70-71 (2001)

COMMENT

COMMENT

Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8659, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanabemange.ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano, S. Construction and Characterization of a Full Length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2) 149-156 (1997).

FEATURES

```

source
1..500
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/isolate="3D7"
/db_xref="taxon:36329"
/clone="XPfM6924"
/dev_stage="erythrocytic stage"
/clone_1ib="Sugano Malaria cDNA library"

```

ORIGIN

```
Query Match      12.4%; Score 41.8; DB 1; length 500;
Best Local Similarity 50.2%; Pred. No. 0.93;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
```

Qy 14 TTAGATAGCCGGTTTCTACTACAACCTCAATAAAGTAGCAGAAATGCGTTAGTGA 73
||| || | | | | | | | | | | | | | | |
Db 232 TTATATTATCATGATATATAGAAAATGGAAATAAAAATAAAAATTAACGAGAAATCATTTAA 29

Db 304 ATAACTGATTCACGAAAGCAGAAATATGTGTCCAAACGGTCAACAATAAGTAAT 245
 QY 95 AAAAGTACTATATCATTTGAGGCATATAGGAGGAGGATTCAGCAACAGTGTGCTT 154
 Db 244 GACAGTTCGAATACCTTTCTGTGTAATAATATGACCCGTGCTTATTTACATTCAGCA 185
 QY 155 ACAAGTGAAGAAACAAGTAACTAAAGTGACCCCTCCCTTGACAATCATATGCACAG 214
 Db 184 CGAAATGGAATAGGTGTTAAATATGTAACATTTAAATTTGTTACGGAATGACACCCCTG 125
 QY 215 TTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTGTAATAGCTGTTCATAA 274
 Db 124 CTAAAGCATCTGCTTATGTGTAATTAATTTCTGTGTTCTTGACAACATAGTTGCAAC 65
 QY 275 TTCTCTTGCAAGCTCTGCTACTTACGAAAGTCTGCTTACAGACAGAAA 325
 Db 64 AGCTTTCTGTAATCTCCACACCTTGCTGCTGCTGCTCTTA 14

RESULT 7
 LOCUS BM140184 730 bp mRNA linear EST 03-NOV-2002
 DEFINITION BM140184 Nori Satoh unpublished cDNA library, gastrula and neurula
 ACCESSION BM140184
 VERSION BM140184.1 GI:24497206
 KEYWORDS EST.
 SOURCE Clona intestinalis
 ORGANISM Clona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Clona.
 1 (bases 1 to 730)
 Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
 Expressed genes in Clona intestinalis (2002c)
 Unpublished (2002)
 JOURNAL Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
 source location/Qualifiers
 1..730
 /organism="Clona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="rcign055e20"
 /issue_type="whole body"
 /dev_stage="gastrula and neurula"
 /clone_1lb="Nori Satoh unpublished cDNA library, gastrula and neurula"

ORIGIN
 Query Match 12.3%; Score 41.4; DB 5; Length 730;
 Best Local Similarity 46.4%; Pred. No. 1.3; Mismatches 156; Indels 0; Gaps 0;
 Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 35 ACACTCAATTAAGATGAACAGATGAGGTGATGACTGTTTAAAGAGACTAAT 94
 Db 40 ATGACTGTATTCAGAAAGCAGAAATATGTGTTCCAAACGGTCAACAATAAGTAAAT 99
 QY 95 AAAGATATCATCATCTTTGAGGCAATTAAGGAGGAGAGATTACCAAAACAGTGTGCTT 154
 Db 100 GACAGTTCGAATACCTTTCTGTAATAATATGACCCGTGCTTATTTACATTTGACCA 159
 QY 155 ACAAGTGAAGAAACAAGTAACTAAAGTGACCCCTCCCTTGACAATCATATGCACAG 214
 Db 160 CGAAATGGAATAGGTGTTAAATATGTAACATTTAAATTTGTTACGGAATGACACCCCTG 219
 QY 215 TTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTGTAATAGCTGTTCATAA 274
 Db 220 CTAAAGCATCTGCTTATGTGTAATTAATTTCTGTGTTCTTGACAACATAGTTGCAAC 279

QY 275 TTCTCTTGCAAGCTCTGCTACTTACAGAAAGTCTGCTTACAGACAGAAA 325
 Db 280 AGCTTTCTGTAATCTCCACACCTTGCTGCTGCTGCTCTTA 330

RESULT 8
 LOCUS BM127139 732 bp mRNA linear EST 02-NOV-2002
 DEFINITION BM127139 Nori Satoh unpublished cDNA library, gastrula and neurula
 ACCESSION BM127139
 VERSION BM127139.1 GI:24483538
 KEYWORDS EST.
 SOURCE Clona intestinalis
 ORGANISM Clona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Clona.
 1 (bases 1 to 732)
 Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
 Expressed genes in Clona intestinalis (2002c)
 Unpublished (2002)
 JOURNAL Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
 source location/Qualifiers
 1..732
 /organism="Clona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="rcign013j16"
 /issue_type="whole body"
 /dev_stage="gastrula and neurula"
 /clone_1lb="Nori Satoh unpublished cDNA library, gastrula and neurula"

ORIGIN
 Query Match 12.3%; Score 41.4; DB 5; Length 732;
 Best Local Similarity 46.4%; Pred. No. 1.3; Mismatches 156; Indels 0; Gaps 0;
 Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 35 ACACTCAATTAAGATGAACAGATGAGGTGATGACTGTTTAAAGAGACTAAT 94
 Db 20 ATAACTGATTCACCAAGCAGAAATATGTGTTCCAAACGGTCAACAATAAGTAAAT 79
 QY 95 AAAGTACTATCATCATCTTTGAGGCAATTAAGGAGGAGAGATTACCAAAACAGTGTGCTT 154
 Db 80 GACAGTTCGAATACCTTTCTGTGTAATAATATGACCCGTGCTTATTTACATTTGACCA 139
 QY 155 ACAAGTGAAGAAACAAGTAACTAAAGTGACCCCTCCCTTGACAATCATATGCACAG 214
 Db 140 CGAAATGGAATAGGTGTTAAATATGTAACATTTAAATTTGTTACGGAATGACACCCCTG 199
 QY 215 TTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTGTAATAGCTGTTCATAA 274
 Db 200 CTAAAGCATCTGCTTATGTGTAATTAATTTCTGTGTTCTTGACAACATAGTTGCAAC 259
 QY 275 TTCTCTTGCAAGCTCTGCTACTTACGAAAGTCTGCTTACAGACAGAAA 325
 Db 260 AGCTTTCTGTAATCTCCACACCTTGCTGCTGCTCTTA 310

RESULT 9
 LOCUS AV877102 759 bp mRNA linear EST 08-NOV-2001
 DEFINITION AV877102 Nori Satoh unpublished cDNA library, tailbud embryo Clona
 ACCESSION AV877102
 VERSION AV877102.1 GI:16864626

KEYWORDS EST.
SOURCE Clona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Clona.
REFERENCE 1 (bases 1 to 759)
AUTHORS Satoh, N., Satou, Y., Kohara, Y. and Shin-i, T.
TITLE Expressed genes in Clona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. 759
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rc1cb31k05"
/issue_type="whole animal"
/dev_stage="tailbud embryo"
/clone_1ib="Nori Satoh unpublished cDNA library, tailbud embryo"

ORIGIN
Query Match 12.3%; Score 41.4; DB 1; Length 759;
Best Local Similarity 46.4%; Pred. No. 1.3;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 35 ACACTCAAAATAGATGAACAGATGAGGTTAGTGAAGCTTTTAAAGAGAGTAAT 94
DB 453 ATAACTGATTCACGAAACAGAAATATGTGTTCCAAACGGTCACAAATAGGTAAT 512
QY 95 AAAGATCATCATCATTTGAGGCATAGAGGAGAGATTCAGCAACAGTGTCTT 154
DB 513 GACAGTCCAAATACCTTTCTGTATAAAATATGACCTGCTTATTCATTGACACA 572
QY 155 ACAAGTGAAAAAAGTTAACTAAAGTACCCCTCTGTCAGCATCAATGACACAG 214
DB 573 CGAATGGAATAGCTTTAAATAGTAAACATTTAAATTTGACGAATGACACCCCTG 632
QY 215 TTGAGCTTAGCCAGCCACATCATCATGTAATTTGCTTTCGTATAGCCTGTTCATAA 274
DB 633 CTAAAGCATCTGCTTATGCTGATTTAATTCGTGTGTTCTTGACACATAGTTGCAAC 692
QY 275 TTCTCTTGCAAGCTCTGCTACTTACCAAGTCTGCTTACAGACAGAAAA 325
DB 693 AGCTTTCTGTAAATCTCCACACATGTCATGCTGCTTCTGCTCTTA 743

RESULT 10
LOCUS BM082303 767 bp mRNA linear EST 22-OCT-2002
DEFINITION BM082303 Nori Satoh unpublished cDNA library, egg Clona intestinalis cDNA clone rclg088c22 3', mRNA sequence.
ACCESSION BM082303
VERSION BM082303.1 GI:24257583
KEYWORDS EST
SOURCE Clona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Clona.
REFERENCE 1 (bases 1 to 767)
AUTHORS Satou, Y., Shin-i, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Clona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. 767
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rclg088c22"
/issue_type="whole animal"
/dev_stage="egg"
/clone_1ib="Nori Satoh unpublished cDNA library, egg"

ORIGIN
Query Match 12.3%; Score 41.4; DB 5; Length 767;
Best Local Similarity 46.4%; Pred. No. 1.3;
Matches 135; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 35 ACACTCAAAATAGATGAACAGATGAGGTTAGTGAAGCTTTTAAAGAGAGTAAT 94
DB 27 ATAACTGATTCACGAAACAGAAATATGTGTTCCAAACGGTCACAAATAGGTAAT 86
QY 95 AAAGATCATCATCATTTGAGGCATAGAGGAGAGATTCAGCAACAGTGTCTT 154
DB 87 GACAGTCCAAATACCTTTCTGTATAAAATATGACCTGCTTATTCATTGACACA 146
QY 155 ACAAGTGAAAAAAGTTAACTAAAGTACCCCTCTGTCAGCATCAATGACACAG 214
DB 147 CGAATGGAATAGCTTTAAATAGTAAACATTTAAATTTGACGAATGACACCCCTG 206
QY 215 TTGAGCTTAGCCAGCCACATCATCATGTAATTTGCTTTCGTATAGCCTGTTCATAA 274
DB 207 CTAAAGCATCTGCTTATGCTGATTTAATTCGTGTGTTCTTGACACATAGTTGCAAC 266
QY 275 TTCTCTTGCAAGCTCTGCTACTTACCAAGTCTGCTTACAGACAGAAAA 325
DB 267 AGCTTTCTGTAAATCTCCACACATGTCATGCTGCTTCTGCTCTTA 317

RESULT 11
LOCUS BM141206 774 bp mRNA linear EST 03-NOV-2002
DEFINITION BM141206 Nori Satoh unpublished cDNA library, gastrula and neurula Clona intestinalis cDNA clone rclgn058h04 3', mRNA sequence.
ACCESSION BM141206
VERSION BM141206.1 GI:24498431
KEYWORDS EST
SOURCE Clona intestinalis
ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Clona.
REFERENCE 1 (bases 1 to 774)
AUTHORS Satou, Y., Shin-i, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Clona intestinalis (2002c)
JOURNAL Unpublished (2002)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1. 774
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="rclgn058h04"
/issue_type="whole body"
/dev_stage="gastrula and neurula"
/clone_1ib="Nori Satoh unpublished cDNA library, gastrula and neurula"

ORIGIN

Query Match 12.3%; Score 41.4; DB 5; Length 774;
 Best Local Similarity 46.4%; Pred. No. 1.3; 156; Indels 0; Gaps 0;
 Matches 135; Conservative 0; Mismatches 0;

QY 35 AACTCAATTAAGATGAAAGATGATGAGTCTGTTATTAAGAAGATTAAT 94
 |||||
 DB 20 ATATGATGATTCACGAAAGACATATGTTCCAAAGGTCACAAATTAAGTAAT 79
 |||||

QY 95 AAAGATCATCATCATCTTTGAGGCAATTAAGGAGGAGGATTCAGCAACAGTGTCTT 154
 |||||
 DB 80 GACAGTTCACATACCTTTCTGTAAAAATAGACCCCTTCATTTTACATTCGACCA 139
 |||||

QY 155 ACAAGTGAAGAAACAACTTAAGTAAAGTACCCCTCCCTGACAAAGATCAATGACAG 214
 |||||
 DB 140 CGAATGGAATAGTGTAAATTAAGTAAACATTAATAATTTACGAAATGACACCCCTG 199
 |||||

QY 215 TTGAGCTTACGACGACATCATCATGTAATGCTTCTGATTAAGCTGTCTATAA 274
 |||||
 DB 200 CTAGCATCTGCTATGCTGATTTAATTTCTGTGTTCTTGAACAATAGTTGCAAC 259
 |||||

QY 275 TTCTCTTGAAGCTCTGCTACTTACGAAAGTCTGCTACACAGACAAA 325
 |||||
 DB 260 AGCTTCTGTAAATCTCCACCATGCTCATGCTCTGCTGCTCTAA 310
 |||||

RESULT 12
 CP250818 598 bp mRNA linear EST 07-AUG-2003
 LOCUS ea017.e05 Elmeiria tenella-infected caecal tonsil Gallus gallus
 DEFINITION CDNA, mRNA sequence.
 ACCESSION CF250818
 VERSION CF250818.1 GI:33484073
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Gallus.
 REFERENCE 1 (bases 1 to 598)
 Witzell, H., Bed Hom, B., Morin, V., Young, J.R., Whitaker, C.J.,
 Chausse, A.M., and Zoorob, R.
 A collection of chicken ESTs from activated immune cells
 Unpublished (2003)
 COMMENT Contact: Zoorob R
 UPR 1983

FEATURES
 source
 1..598
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 /clone_idb="Elmeiria tenella-infected caecal tonsil"
 /note="Organ: Caecal tonsil; Vector: pTribLEX2"

ORIGIN
 Query Match 12.2%; Score 41; DB 7; Length 598;
 Best Local Similarity 100.0%; Pred. No. 1.6;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 ACTTACCAAGAGTCTGCTTACAGACAGAAAGATGGATTTA 336
 |||||
 DB 3 ACTTACCAAGAGTCTGCTTACAGACAGAAAGATGGATTTA 43
 |||||

RESULT 13
 AZ305625 594 bp DNA linear GSS 29-SEP-2000
 LOCUS AZ305625/c
 DEFINITION IM0006F20F Mouse 10kb plasmid UUCGIM library Mus musculus genomic

clone UUCGIM0006F20 F, genomic survey sequence.
 AZ305625
 AZ305625.1 GI:10342825
 GSS.
 KEYWORDS Mus musculus
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 594)
 Authors
 Dun, D., Aoyagi, A., Barber, M., Beecorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
 Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
 Niederhausern, A., and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 JOURNAL
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddun@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0006 row: P column: 20
 Seq primer: CGTTGTAAACGACGCGCACT
 Class: plasmid ends
 High quality sequence stop: 594.
 Location/Qualifiers
 1..594
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCGIM0006F20"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /clone_idb="Mouse 10kb plasmid UUCGIM library"
 /note="Vector: pMD42nv, Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN
 Query Match 12.1%; Score 40.6; DB 8; Length 594;
 Best Local Similarity 53.5%; Pred. No. 2.1; 74; Indels 0; Gaps 0;
 Matches 85; Conservative 0; Mismatches 74;

QY 22 CCGGTTTTTACTCAACTCAATTAAGATGAAAGATGATGAGTCTGTTAT 81
 |||||
 DB 508 CTGATTATTATTCTCAAGAGAAAGAAAGAAAGAAAGTCAAGAAAGAA 449
 |||||

QY 82 AAAGAAAGATTAATAAGATCATCTATCTTTGAGGCAATTAAGGAGGAGAGATTTCAC 141
 |||||
 DB 448 AAAGAAAGATGAAGAAAGAAAGAAAGAAAGGAGGAGGAGGAGGAGGAGGAGG 389
 |||||

QY 142 AAACAGTGTGCTTACAGTGAAGAAACAGTTAACTAAA 180
 |||||

This Page Blank (uspto)

CC into a host cell (and comprising the isolated avian gut-specific gene
CC expression control region), expressing a heterologous polypeptide in a
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA
CC molecule, and culturing the transfected cell in a medium suitable for
CC expression of a heterologous polypeptide under the control of an avian
CC intestinal fatty acid binding protein (IFABP) or cp35 gene expression
CC control region encoded by the recombinant DNA molecule), a eukaryotic
CC cell transformed with the expression vector (or its progeny, which
CC expresses a heterologous polypeptide) and a transgenic avian having a
CC heterologous polynucleotide sequence comprising the nucleic acid insert.
CC The nucleic acids are useful for regulating heterologous nucleic acids in
CC transgenic avians, as probes in nucleic acid hybridisation assays for
CC detecting the IFABP gene expression control region, and for generating
CC transgenic birds. The present sequence is the Chicken intestinal fatty
CC acid binding protein, IFABP, gene, promoter.

XX Sequence 336 BP; 116 A; 62 C; 63 G; 95 T; 0 U; 0 Other;

SO Query Match 100.0%; Score 336; DB 11; Length 336;

Best Local Similarity 100.0%; Pred. No. 2e-87; Mismatches 0; Gaps 0;

Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTTATTTTCATTAGATAGCCGGTTTCTTACTACAACTCAATTAAGTGAACAGAAATG 60
DB 1 ATTATTTATTTTCATTAGATAGCCGGTTTCTTACTACAACTCAATTAAGTGAACAGAAATG 60
QY 61 AATGGTTTATGTCCTTTTAAAGAGTAAATGATCTATCTATCTTTTGGGCA 120
DB 61 AATGGTTTATGTCCTTTTAAAGAGTAAATGATCTATCTATCTTTTGGGCA 120
QY 121 TAAAGGAGGAGAGATTTCAGCAAAACAGTGTCTTACAGTGAAGAAACAAGTTAACTAA 180
DB 121 TAAAGGAGGAGAGATTTCAGCAAAACAGTGTCTTACAGTGAAGAAACAAGTTAACTAA 180
QY 181 GTGACCCCCCTCTTGACAAAGATCAATGCCACAGTTGAGCTTGGCCACCATCATCA 240
DB 181 GTGACCCCCCTCTTGACAAAGATCAATGCCACAGTTGAGCTTGGCCACCATCATCA 240
QY 241 TGTAAATGCTTCCGATTAAGCTCTTCAATAATCTCTTTGCAAAAGCTGTGACTTA 300
DB 241 TGTAAATGCTTCCGATTAAGCTCTTCAATAATCTCTTTGCAAAAGCTGTGACTTA 300
QY 301 CCAGAAGTCTGCTACAGACAGAAAGATGGCAATTTA 336
DB 301 CCAGAAGTCTGCTACAGACAGAAAGATGGCAATTTA 336

RESULT 2
ADL90127
ID ADL90127 standard; DNA; 2381 BP.

AC ADL90127;

DT 20-MAY-2004 (first entry)

DE Chicken intestinal fatty acid binding protein, IFABP, gene, 5' region.

XX Chicken; de; intestinal fatty acid binding protein; IFABP;

KW gut specific promoter; transgenic.

OS Gallus gallus.

XX US2003177516-A1.

XX 18-SEP-2003.

XX 14-MAR-2002; 2002US-00099663.

XX 14-MAR-2002; 2002US-00099663.

XX (HORS/) HORSEMAN N D.

XX (PRAT/) PRATT S L.

XX

PI Horseman ND, Pratt SL;
XX WPI; 2003-898653/82.

XX New nucleic acid molecule comprising an isolated avian gut-specific gene
XX expression control region, useful for regulating heterologous nucleic
XX acids in transgenic avians, and for generating transgenic birds.

PS Claim 1; SEQ ID NO 1; 28bp; English.

XX The invention relates to an isolated nucleic acid comprising an isolated
XX avian gut-specific gene expression control region appearing as
XX ADL90127 (Chicken intestinal fatty acid binding protein, IFABP, gene, 5'
XX region or ADL90128 (chicken IFABP promoter) or its degenerate variant.
XX Also included are a recombinant DNA molecule comprising an isolate avian
XX gut-specific gene expression control region operably linked to a nucleic
XX acid insert encoding a polypeptide, an expression vector that integrates
XX into a host cell (and comprising the isolated avian gut-specific gene
XX expression control region), expressing a heterologous polypeptide in a
XX host cell (by transfecting a eukaryotic cell with the recombinant DNA
XX molecule, and culturing the transfected cell in a medium suitable for
XX expression of a heterologous polypeptide under the control of an avian
XX intestinal fatty acid binding protein (IFABP) or cp35 gene expression
XX control region encoded by the recombinant DNA molecule), a eukaryotic
XX cell transformed with the expression vector (or its progeny, which
XX expresses a heterologous polypeptide) and a transgenic avian having a
XX heterologous polynucleotide sequence comprising the nucleic acid insert.
XX The nucleic acids are useful for regulating heterologous nucleic acids in
XX transgenic avians, as probes in nucleic acid hybridisation assays for
XX detecting the IFABP gene expression control region, and for generating
XX transgenic birds. The present sequence is the Chicken intestinal fatty
XX acid binding protein, IFABP, gene, 5' region.

XX Sequence 2381 BP; 775 A; 428 C; 469 G; 709 T; 0 U; 0 Other;

SO Query Match 100.0%; Score 336; DB 11; Length 2381;

Best Local Similarity 100.0%; Pred. No. 3.8e-87; Mismatches 0; Gaps 0;

Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTTATTTTCATTAGATAGCCGGTTTCTTACTACAACTCAATTAAGTGAACAGAAATG 60
DB 1301 ATTATTTATTTTCATTAGATAGCCGGTTTCTTACTACAACTCAATTAAGTGAACAGAAATG 1360
QY 61 AATGGTTTATGTCCTTTTAAAGAGTAAATGATCTATCTATCTTTTGGGCA 120
DB 1361 AATGGTTTATGTCCTTTTAAAGAGTAAATGATCTATCTATCTATCTTTTGGGCA 1420
QY 121 TAAAGGAGGAGAGATTTCAGCAAAACAGTGTCTTACAGTGAAGAAACAAGTTAACTAA 180
DB 1421 TAAAGGAGGAGAGATTTCAGCAAAACAGTGTCTTACAGTGAAGAAACAAGTTAACTAA 1480
QY 181 GTGACCCCCCTCTTGACAAAGATCAATGCCACAGTTGAGCTTGGCCACCATCATCA 240
DB 1481 GTGACCCCCCTCTTGACAAAGATCAATGCCACAGTTGAGCTTGGCCACCATCATCA 1540
QY 241 TGTAAATGCTTCCGATTAAGCTCTTCAATAATCTCTTTGCAAAAGCTGTGACTTA 300
DB 1541 TGTAAATGCTTCCGATTAAGCTCTTCAATAATCTCTTTGCAAAAGCTGTGACTTA 1600
QY 301 CCAGAAGTCTGCTACAGACAGAAAGATGGCAATTTA 336
DB 1601 CCAGAAGTCTGCTACAGACAGAAAGATGGCAATTTA 1636

RESULT 3
ABL18616

ID ABL18616 standard; DNA; 2933 BP.

AC ABL18616;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7321.

XX	Drosophila; developmental biology; cell signalling; insecticide;
KM	pharmaceutical; gene; db.
XX	
OS	Drosophila melanogaster.
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
PR	11-UL-2000; 2000US-00614150.
XX	
PA	(PEXE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions.
PS	
XX	Claim 1; SEQ ID NO 7321; 21bp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC	sequences (AB101840-AB16175) and the encoded proteins (AB57737-
CC	AB572072). The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SO	Sequence 2933 BP; 883 A; 628 C; 606 G; 816 T; 0 U; 0 Other;
XX	
Query Match	11.6%; Score 39; DB 4; Length 2933;
Best Local Similarity	52.1%; Pred. No. 0.75;
Matches 87; Conservative	0; Mismatches 80; Indels 0; Gaps 0;
OY	73 ACTGTTTAAAGAGAGTAATTAAGATCATCATCTTGAAGCAATTAAGGAGGAG 132
DB	2766 ACATTAAATGCACGATCTTACTTAAAGAAATATAGATTATGATGACAGAAAGAGAT 2825
OY	133 AGATTACGAAACAGCTGTGCTTACAAAGTGAAGAAACCACTTAAACTTAAGTACCCCTC 192
DB	2826 ATATTAAACCAAAAATTTCTTATCGATGTGAACCAAGTGAACCGAAATTTGCATTTC 2885
OY	193 CTGACAAAGATCAATGCCACAGTTGAGCTTAAAGCCAGCAATCATC 239
DB	2886 CGTGCCCACTGAATGCAATCACTTAATGCAACACCAACCAATCAACC 2932
RESULT 4	
ABL06564	
ID	ABL06564 standard; cDNA; 2927 BP.
XX	
AC	ABL06564;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 1417A.
XX	
KM	Drosophila; developmental biology; cell signalling; insecticide;
KM	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.

XX	PD	27-SEP-2001.
XX	PF	23-MAR-2001; 2001WO-US009231.
XX	PR	23-MAR-2000; 2000US-0191637P.
XX	PR	11-JUL-2000; 2000US-00614150.
XX	PA	(PEKE) PE CORP NY.
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	DR	WPI; 2001-656860/75.
XX	DR	P-PsDB; ABB62461.
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX	PT	
XX	PS	Claim 1; SEQ ID NO 14174; 21pp + Sequence Listing; English.
XX	CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	CC	
XX	SQ	Sequence 2927 BP; 882 A; 623 C; 606 G; 816 T; 0 U; 0 Other;
XX		
XX	Query March	11.1%; Score 37.2; DB 4; Length 2927;
XX	Best Local Similarity	51.9%; Pred. No. 2.5;
XX	Matches	84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
XX	Dy	73 ACTGTTATAAAGAGAAGTATATAAGATTACATCATCTTGAGCAATAAGAGAGAG 132
XX	Dd	2766 ACATTAAATGCAGCATCTTAATTACGAATTAATGATTAATGTCGAGMAAGAGAGAT 2825
XX	Dy	133 AGATTGACGAACAAGTGTGCTTACAGGTGAAAAACAAGTTAAACTAAATGAGACCCCCTC 192
XX	Dd	2826 ATATTAAACGAATAATTCTTATCGATGTGAACAGTAGAACGAAATTTGCATTCATC 2885
XX	Dy	193 CTGACAAAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACA 234
XX	Dd	2886 CGTGCCCCACTGATGCATCATCACTTAATGCGACACCACCATTA 2927
XX		
XX	RESULT 5	
XX	ID ADJ79962/c	
XX	ID ADJ79962 standard; DNA; 123526 BP.	
XX	ADJ79962;	
XX	DT 06-MAY-2004 (first entry)	
XX	DE Human glioma-associated oncogene-3 related DNA, SEQ ID No 11.	
XX	KW glioma-associated oncogene-3; GAO3; cytosolic; developmental disorder; Greig's cephalopolysyndactyl; Pallister-Hall syndrome;	
XX	KM post-axial polydactyl; holoprosencephaly; Rubenstein-Teybi syndrome; basal cell nevold syndrome; hyperproliferative disorder; cancer; human; ds.	
XX	KM	
XX	OS Homo sapiens.	
XX	PN WO2003008549-A2.	
XX	PD 30-JAN-2003.	
XX		

XX	(CORI-) CORIXA CORP.
PA	
XX	
PI	Meagher MJ, Xu J, King GB;
XX	
DR	WPI; 2001-611627/70.
XX	
FT	New colon tumor proteins and related nucleic acid, useful for treatment,
PT	prevention, diagnosis and monitoring of cancer.
XX	
PS	Claim 4; Page 202; 299pp; English.
XX	
CC	The present invention relates to the isolation of novel cDNA sequences
CC	encoding for at least an immunogenic portion of human colon tumour
CC	proteins. The sequences of the invention are useful in pharmaceutical
CC	compositions and vaccines for the prevention and treatment of cancers
CC	such as colon cancer. They are also useful for the diagnosis and
CC	monitoring of such cancers. Antibodies to the colon tumour proteins and
CC	antigen presenting cells that express polynucleotides encoding colon
CC	tumour proteins can be used to inhibit the development of cancers. T-
CC	cells that react specifically with colon tumour proteins are useful for
CC	removing tumour cells from samples (e.g. blood) and for cancer treatment.
CC	The polynucleotide sequences are also useful in gene therapy. AAS57325-
CC	AAS5880 represent the cDNA sequences of the invention that encode for
CC	portions of human colon tumour proteins
XX	
SQ	Sequence 403 BP; 136 A; 63 C; 88 G; 105 T; 0 U; 11 Other;
Query Match	10.9%; Score 36.6; DB 4; Length 403;
Best Local Similarity	50.6%; Pred No.1.9; 82; Indels 0; Gaps 0
Matches	84; Conservative 0; Mismatches 0; Gaps 0
Qy	26 TTTTTCACACAACTCAATTAAGATGACAGATGAATGGTTAGTGTATTATTAAG 85
Db	178 TGTTTAAAGTTAAAGAAAAAAGAGCTGACAGATATTATTAATCTGCTTTTAGAAAA 237
Qy	86 AAGAGTAATTAAGATCTATCATCTTTGAGGCAATTAAGGAGGAGAGATTACGCAAC 145
Db	238 AACAAACCAAGAAAGACCATTTGACCATTAATGAAAAAGGAAAGATTTATATAGAAC 297
Qy	146 AGCTGCTTCAAGTGAAGAAAAAACAATTAACTTAAGTACCCCT 191
Db	298 TTGTCTAGTTNAAAAAAAAAAAAAAAACTTNGCNGAACCCCT 343
RESULT 7	
ABLI0846	
ID	ABLI0846 standard; cDNA; 3322 BP.
XX	
AC	ABLI0846;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 27020.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical; gene; ss.
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
XX	
XX	11-JUL-2000; 2000US-00614150.
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PMD, Myers EW;
XX	

DR WPI: 2001-656860/75.
P-PSDB; ABB66743.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from *Drosophila* and for elucidating cell signaling and cell-cell
interactions.
XX
PS Claim 1; SEQ ID NO 27020; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from *Drosophila*. The invention is
useful in developmental biology and in elucidating cell signaling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
sequences (AB101840-AB16175) and the encoded proteins (AB57737-
AB872072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at http://wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3322 BP; 993 A; 715 C; 742 G; 872 T; 0 U; 0 Other;
Query Match 10.9%; Score 36.6; DB 4; Length 3322;
Best Local Similarity 57.4%; Pred. No. 3.9;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 1 ATTTATTTTTCATTTAGATACCGCGTTTCTACTACACTCAATAAGATGAAGATG 60
DB 3191 ATTGAGAAATTTCTTTTATCTGTTTGTATATCTACACAATTTGGTCCAGTAGA 3250
QY 61 AATGGGTAGTGCCTGTTTAAAGAGAGTAATTAAGATCATCTATTGGA 115
DB 3251 AGTTAGATTTTAATCGGTAGCAGAGTCAATTAACAATTAATTAATCTTGA 3305
RESULT 8
AA520251/C
ID AA520251 standard; DNA; 53585 BP.
AC AA520251;
XX
DT 04-MAY-1999 (first entry)
XX
DE *Borrelia burgdorferi* polynucleotide sequence #4.
XX
KW *Borrelia burgdorferi*; spirochete; bacterium; pathogen; Lyme disease;
KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX infection; diagnosis; characterisation; detection; ds.
OS *Borrelia burgdorferi*.
XX
PN WO9858943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US012764.
XX
PR 20-JUN-1997; 97US-0050359P.
XX 22-JUL-1997; 97US-0053344P.
PR 22-JUL-1997; 97US-0053377P.
PR 03-SEP-1997; 97US-0057483P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (MED-) MEDIMUNE INC.
XX
PI Frazer C, White OR, Clayton R, Dougherty BA, Lathigra R;
PI Smith HO;
XX
DR WPI, 1999-081217/07.
XX
PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention and
PT therapy of infections, particularly Lyme disease.
XX

XX
PS Claim 1; Page 801-831; 1128bp; English.
XX
CC AA520248 to AA520402 represent polynucleotide sequences isolated from
CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for the
CC detection, diagnosis, characterisation, prevention and therapy of Bb
CC infections, e.g. Lyme disease. They can also be used for the production
CC of biosynthetic products, e.g. enzymes. *Borrelia* belongs to a family of
CC motile, spiral-shaped bacteria called Spirochetes. Spirochetes are
CC pathogenic in humans and *Borrelia* causes epidemic and endemic relapsing
XX fever, and Lyme borreliosis, more commonly known as Lyme disease
XX
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T; 0 U; 2 Other;
Query Match 10.9%; Score 36.6; DB 2; Length 53585;
Best Local Similarity 58.9%; Pred. No. 10;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 78 TTATTAAGAAGAGATTAATAAGATCTATCTCTTTGAGGCATTAAGGAGAGAGATT 137
DB 44978 TTTAAAAATGTTATTAATAAGATTACTCACTTATATCAAGCATTAATGATGAGAGAGATG 44919
QY 138 CAGCAAAACGTGCTTACAGTGAAGAAACAGTTAACTAAGTGA 184
DB 44918 CAGCATCTATTGAACATTTTAAGAGAGAGAGAAATGAACTAATGTCA 44872
RESULT 9
AAH53268
ID AAH53268 standard; DNA; 1587 BP.
AC AAH53268;
XX
DT 03-SEP-2001 (first entry)
XX
DE *S. epidermidis* open reading frame nucleotide sequence SEQ ID NO:1929.
XX
KW *Staphylococcus epidermidis* SRI strain; infection; diagnosis; vaccination;
XX endocarditis; ds.
XX
OS *Staphylococcus epidermidis*.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAX) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
DR WPI, 2001-316495/33.
DR P-PSDB; AAG82418.
XX
PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
PT useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 527-528; 2188bp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the *S.*
CC *epidermidis* polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to

AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464

XX Sequence 1587 BP; 617 A; 268 C; 201 G; 501 T; 0 U; 0 Other;

Query Match 10.8%; Score 36.2; DB 4; Length 1587;
Best Local Similarity 53.1%; Pred. No. 3.9;

Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTCAAGTGAACAAAGTTAACTAAAGACCCCTCTTGACAA 200

DB 203 CAGACGATGACTTTAAAGTTGAACCTTATTTTACAGAGTTACTCAAGCCCATACTA 262

QY 201 GATCAATGCGACAGTTAGCTTACGACCATCATCATGTAATGCTTCTTGATA 260

DB 263 GAAATATTAACAAATTAATCTCAAGCATATAGATATCATCATTAATCACTTGAAA 322

QY 261 AGCCTGTTCAATAATTCCTTTGCA 285

DB 323 ATAAATACACAAATTACTTGCA 347

RESULT 10

ABN92204 standard; DNA; 1878 BP.

XX ABN92204;

DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1667.

XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

XX antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.

XX 14-AUG-1997; 97US-0055779P.

XX 08-NOV-1997; 97US-0064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX P-PSDB; ABP93659.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermis

XX PT polypeptide, useful for diagnosing and treating bacterial infections.

XX Disclosure; SEQ ID NO 1667; 267bp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading

XX CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

XX CC given in ABP35124 to ABP37960. The S. epidermidis sequences have

XX CC antibacterial activity and can be used in gene therapy. The sequences can

XX CC also be used in the diagnosis and treatment of bacterial infections,

XX CC particularly S. epidermidis infections. The sequences can be used to

XX CC screen for compounds able to interfere with the S. epidermidis life cycle

XX CC or inhibit S. epidermidis infection. N.B. The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from the USPTO web site

XX Sequence 1878 BP; 734 A; 304 C; 239 G; 601 T; 0 U; 0 Other;

XX Query Match 10.8%; Score 36.2; DB 6; Length 1878;
Best Local Similarity 53.1%; Pred. No. 4.2;

Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGTGTGCTTCAAGTGAACAAAGTTAACTAAAGACCCCTCTTGACAA 200

DB 494 CAGACGATGACTTTAAAGTTGAACCTTATTTTCAAGAGTTACTCAAGCCCATACTA 553

QY 201 GATCAATGCGACAGTTAGCTTACGACCATCATCATGTAATGCTTCTTGATA 260

DB 554 GAAATATTAACAAATTAATCTCAAGCATATAGATATCATCATTAATCACTTGAAA 613

QY 261 AGCCTGTTCAATAATTCCTTTGCA 285

DB 614 ATAAATACACAAATTACTTGCA 638

RESULT 11

ADSO1975 standard; DNA; 1878 BP.

XX ADSO1975;

DT 04-NOV-2004 (first entry)

XX Staphylococcus epidermis polynucleotide seqid 1270.

XX antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;

XX recombinant expression vector; infection; computer readable medium;

XX computer based system; gene; ds.

XX Staphylococcus epidermidis.

XX US2004147734-A1.

XX 29-JUL-2004.

XX 01-DEC-2003; 2003US-00724972.

XX 08-NOV-1997; 97US-0064964P.

XX 13-AUG-1998; 98US-00134001.

XX 29-NOV-1999; 99US-00450969.

XX (DOUC/) DOUCETTE-STAMM L.

XX (BUSH/) BUSH D.

XX Doucette-Stamm L, Bush D;

XX WPI; 2004-580138/56.

XX P-PSDB; ADS05747.

XX New isolated polypeptide and encoding nucleic acid derived from

XX PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or

XX treating an S. epidermidis bacterial infection.

XX Claim 5; SEQ ID NO 1270; 741bp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide

XX CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:

XX CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any

XX CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as

XX CC given in the specification. Also described are: a recombinant expression

XX CC vector; a cell comprising a recombinant expression vector of (1);

XX CC producing an S. epidermidis polypeptide; an isolated nucleic acid

XX CC comprising a nucleotide sequence of at least 8 nucleotides in length; a

XX CC vaccine composition for prevention or treatment of an S. epidermidis

XX CC infection, comprising a nucleic acid cited above and a carrier; treating

XX CC a subject for S. epidermidis infection; a recombinant or substantially

CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This sequence encodes a S. epidermidis protein of the invention.
SQ Sequence 1878 BP; 734 A; 304 C; 239 G; 601 T; 0 U; 0 Other;
Query Match 10.8%; Score 36.2; DB 13; Length 1878;
Best Local Similarity 53.1%; Pred. No. 4.2; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 141 CAACAGTGTGCTTACAGTGAAGAAACAACTTAAGTAAAGTACCCCTCTTGACAA 200
DB 494 CAGACGATGACTTTAAAGTTGAAAACCTTATTTACAGAGTTACCTCAAGCCATACTA 553
QY 201 GATCAATGCCACAGTTCAGCTTTAGCCAGCCATCATCATGTAATTGCTTCTGATA 260
DB 554 GAAAAATTAACAAATATATCTCAAGCATATAGATATATCATTAATCAATCTTGAAA 613
QY 261 AGCCTGTTCATTAATTCCTTTGCA 285
DB 614 ATAAATATACAAATTAATTCCTTTGCA 638
RESULT 12
AAH54543
ID AAH54543 standard; DNA; 3000 BP.
AC AAH54543;
XX
XX
DT 03-SEP-2001 (first entry)
XX
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3907.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
PT WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 1560-1561; 2188BP; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAH51454 to AAH53120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing

CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
SQ Sequence 3000 BP; 1019 A; 503 C; 512 G; 966 T; 0 U; 0 Other;
Query Match 10.8%; Score 36.2; DB 4; Length 3000;
Best Local Similarity 53.1%; Pred. No. 4.9;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 141 CAACAGTGTGCTTACAGTGAAGAAACAACTTAAGTAAAGTACCCCTCTTGACAA 200
DB 129 CAGACGATGACTTTAAAGTTGAAAACCTTATTTACAGAGTTACCTCAAGCCATACTA 188
QY 201 GATCAATGCCACAGTTCAGCTTTAGCCAGCCATCATCATGTAATTGCTTCTGATA 260
DB 189 GAAAAATTAACAAATATATCTCAAGCATATAGATATATCATTAATCAATCTTGAAA 248
QY 261 AGCCTGTTCATTAATTCCTTTGCA 285
DB 249 ATAAATATACAAATTAATTCCTTTGCA 273
RESULT 13
AAH54805/C
ID AAH54805 standard; DNA; 3368 BP.
AC AAH54805;
XX
XX
DT 03-SEP-2001 (first entry)
XX
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4169.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US030782.
XX
PR 09-NOV-1999; 99US-0164258P.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Kimerly WJ;
XX
PT WPI; 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis.
XX
PS Claim 8; Page 1875-1876; 2188BP; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAH51454 to AAH53120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.

CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464

XX SQ Sequence 3368 BP; 1163 A; 447 C; 569 G; 1189 T; 0 U; 0 Other;

XX Query Match 10.8%; Score 36.2; DB 4; Length 3368;
XX Best Local Similarity 53.1%; Pred. No. 5.1;
XX Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAACAGCTGTCTTACAGAGTGAACCAAGTTAAAGTACCCCTCTTGACAA 200
DB 1038 CAGACCATGACTTAAAGTTGAACCTTATTACAGAGTTACCTCAAGCCATACTA 979

QY 201 GATCATGCGACAGCTTACGAGTACGACCATATCATCATGTAATGCTTCCGATA 260
DB 978 GAAATATTAACCAATATCTCAAGCATATGATATCATCATTAATTAACCTTGAA 919

QY 261 AGCCTGTCTAATATCTCTTGTGA 285
DB 918 ATTAATATCAACAATTAATTCTGTGCA 894

XX Db

XX RESULT 14
XX AAH84331
XX ID AAH84331 standard; DNA; 6088 BP.

XX AC AAH84331;
XX DT 08-SEP-1999 (first entry)

XX DE Stealth virus nucleic acid clone, SEQ ID NO: 23.

XX KM Stealth virus; detection; diagnosis; infection; ss.

XX OS Stealth virus.

XX FH Key Location/Qualifiers
XX FT misc_difference 5841
XX FT /*tag= a
XX FT /note= "this nucleotide is represented as a * in the
XX FT specification, and is included to maintain the base
XX FT numbering given in the specification"
XX FT 5997. .8133
XX FT /*tag= b
XX FT /note= "this nucleotide is represented as a * in the
XX FT specification, and is included to maintain the base
XX FT numbering given in the specification"
XX FT misc_difference 8275
XX FT /*tag= d
XX FT /note= "this nucleotide is represented as a * in the
XX FT specification, and is included to maintain the base
XX FT numbering given in the specification"

XX PN WO9934019-A1.
XX PD 08-JUL-1999.
XX PF 30-DEC-1998; 98WO-US027744.
XX XX

PR 30-DEC-1997; 97US-00001184.
XX XX
XX PA (MART/) MARTIN W J.
XX PI Martin WJ;
XX DR WPI; 1999-405521/34.
XX PT Novel strains of stealth virus.
XX PS Claim 19; Page 64-66; 95pp; English.

XX CC This sequence represents a Stealth virus nucleic acid clone. The
XX CC invention relates to a method of detecting and characterising a stealth
XX CC virus by reacting a sample suspected of containing a stealth virus with a
XX CC probe from a known stealth virus and sequencing the resultant isolated
XX CC nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
XX CC from a sample suspected of containing a stealth virus, e.g. a culture of
XX CC cells showing a viral cytopathic effect; (b) testing the reactivity of
XX CC the isolated DNA or RNA with a molecular probe that contains at least 18
XX CC or more contiguous nucleotides identical to sequence previously
XX CC identified from a stealth virus; and, optionally (c) sequencing the
XX CC isolated DNA or RNA molecules that react with the probe. The method is
XX CC used to detect stealth virus in a biological product, food or in the
XX CC environment. The method is also used to evaluate agents for their
XX CC inhibitory or stimulatory effects on stealth virus replication and to
XX CC determine capacity of the virus to recombine with and potentially alter
XX CC the nucleic acid sequences of a cell or bacterium

XX SQ Sequence 6088 BP; 1577 A; 1455 C; 1442 G; 1427 T; 0 U; 187 Other;

XX Query Match 10.8%; Score 36.2; DB 2; Length 6088;
XX Best Local Similarity 20.1%; Pred. No. 6.3;
XX Matches 43; Conservative 77; Mismatches 94; Indels 0; Gaps 0;

QY 42 AATTAAGATGACAGATGATGAGTGTAGTACTGTTTATTAAGAAGTATTAAGATA 101
DB 5824 AATTGAAGAATCTTCHANYWVTSRNMNMAMHADATAGRAHNRNMNATNATNTST 5883

QY 102 CTATCATCTTTGAGGCAATTAAGGAGGAGAGATTCAGCAACAGTGTCTTACAAGTG 161
DB 5884 ATHBDCMBGTSMNRMANSYVBARAHCBHNMARTNMHNRTHDITBCEHNMARTN 5943

QY 162 GAAACCAAGTTAACTAAGTACCCCTCTTGACAGATCAATGACAGCTGAGCT 221
DB 5944 NEMAMHNMARTMCGSTYRDMMDHCCDBCTTGDANBCDAADRTNRYTABNRDDCM 6003

QY 222 TTAGCCAGCCATCATCATGTAAATGCTTTC 255
DB 6004 NTRSMWARYNRWATNDCMWTSMWARYNRWATNBC 6037

XX Db

XX RESULT 15
XX ABK31354
XX ID ABK31354 standard; DNA; 8648 BP.

XX AC ABK31354;
XX DT 23-APR-2002 (first entry)

XX DE Signal transduction associated gene modified DNA #99.

XX KM Human; signal transduction associated gene; cytosine methylation state;
XX KM CpG island; signal transduction associated disease; solid tumour; cancer;
XX KM antitumour; cytosolic; mutant; de.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200200926-A2.
XX PD 03-JAN-2002.
XX XX

PF 29-JUN-2001; 2001WO-EP007472.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

XX MPI; 2002-147896/19.

XX

PT Oligonucleotide for diagnosis and therapy of diseases associated with

PT signal transduction e.g. cancer, comprises chemically modified genomic

PT sequences of genes associated with signal transduction.

XX

PS Claim 1; SEQ ID NO 197; 24pp; English.

XX

CC The present invention relates to chemically modified DNA sequences of

CC signal transduction associated genes. The DNA sequences are chemically

CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.

CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting

CC the cytosine methylation state (Cpg islands) of these genes, and a method

CC for the diagnosis and/or therapy of genetic and epigenetic parameters of

CC genes associated with signal transduction. The genomic DNA can be

CC obtained from cells or cellular components which contain DNA, e.g. cell

CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,

CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,

CC brain, heart, prostate, lung, breast or liver, histologic object slides,

CC and all their possible combinations. The sequences of the invention are

CC useful for the diagnosis and therapy of diseases associated with signal

CC transduction e.g. solid tumours and cancer. ABK3158-ABK31545 represent

CC chemically pretreated genomic DNA sequences of different genes associated

CC with signal transduction, or their complementary sequences. Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

XX European Patent Office

XX

SQ Sequence 8648 BP; 2149 A; 236 C; 1917 G; 4346 T; 0 U; 0 Other;

Query Match 10.8%; Score 36.2; DB 6; Length 8648;

Best Local Similarity 57.5%; Pred. No. 7.1;

Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 3 TATTTATTTTCATTAGTACCGGTTTTTTTACTACAACTCAATAGATGACGAAATGAA 62

DB 940 TTTTGGTTTATTACGTGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 999

QY 63 TGGGTTAGTACCTGTTTATTAAGAAAGTAAATTAAGTACTATCATCATTTGA 115

DB 1000 TTGGGAAGTGAATGTTTATGAGAGTAAATAGATTATATTTTGTATTATGA 1052

Search completed: November 21, 2005, 05:53:48
Job time : 271.264 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2005, 02:06:40 ; Search time 1583.72 Seconds
(without alignments)
10280.216 Million cell updates/sec

Title: US-10-099-663-2

Perfect score: 336

Sequence: 1 attattatttcattcattagata.....agacagaaagatgcattta 336

Scoring table: IDENTIFY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sta:*
12: gb_sy:*
13: gb_un:*
14: gb_v1:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	42.8	12.7	88688	9	AC087429	AC087429 Homo sapi
C 2	42.8	12.7	187003	2	AC023353	AC023353 Homo sapi
C 3	41.8	12.4	252420	3	AE014841	AE014841 Plasmid
C 4	41.4	12.3	1663	3	AK116155	AK116155 Clona int
C 5	41.4	12.3	99855	8	OSJN00175	OSJN00175 Oryza sat
C 6	41.4	12.3	121130	8	OSJN00292	OSJN00292 Oryza sat
C 7	40.6	12.1	209024	10	AL928605	AL928605 Mouse DNA
C 8	40.4	12.0	60756	2	AL583823_3	AL583823_3 Human DNA
C 9	40.4	12.0	131708	2	AL354710	AL354710 Homo sapi
C 10	40.4	12.0	174294	2	AC026155	AC026155 Homo sapi
C 11	40.2	12.0	110000	10	AE014175_2	AE014175_2 Mus muscu
C 12	40.2	12.0	201783	10	AC114410	AC114410 Mus muscu
C 13	39.8	11.8	177655	2	CR318618	CR318618 Danio rer
C 14	39.8	11.8	217375	5	BX537337	BX537337 Zebrafish
C 15	39.6	11.8	162605	5	AC068698	AC068698 Homo sapi
C 16	39.6	11.8	164275	9	AC005157	AC005157 Homo sapi
C 17	39.6	11.8	230585	10	AC110259	AC110259 Mus muscu
C 18	39.4	11.7	393	8	YSCMTORIE	K02992 Yeast (S.ce
C 19	39.4	11.7	396	8	MIECORIS	X01216 Yeast mtoc

C 20	39.4	11.7	130574	9	AL160174	AL160174 Human DNA
C 21	39.4	11.7	141668	10	BX001028	BX001028 Mouse DNA
C 22	39.4	11.7	172945	9	AC067719	AC067719 Homo sapi
C 23	39.4	11.7	193306	2	AC025512	AC025512 Homo sapi
C 24	39.4	11.7	194638	2	AC112913	AC112913 Homo sapi
C 25	39.4	11.7	238301	2	AC094376	AC094376 Rattus no
C 26	39	11.6	2933	6	CQ597407	CQ597407 Sequence
C 27	39	11.6	9124	2	AC020330	AC020330 Drosophila
C 28	39	11.6	86258	5	BX842699	BX842699 Zebrafish
C 29	39	11.6	108924	3	AC005889	AC005889 Drosophila
C 30	39	11.6	177992	3	AC092492	AC092492 Drosophila
C 31	39	11.6	270775	3	AE003624	AE003624 Drosophila
C 32	38.8	11.5	1031	3	TERHIS2ALA	TERHIS2ALA Tetrahymena
C 33	38.8	11.5	109465	8	AP006106	AP006106 Lotus cor
C 34	38.8	11.5	140266	8	AC124406	AC124406 Mus muscu
C 35	38.8	11.5	210359	9	HS4430K20	HS4430K20 Human DNA
C 36	38.8	11.5	212445	10	AC108419	AC108419 Mus muscu
C 37	38.6	11.5	165873	5	AL954323	AL954323 Zebrafish
C 38	38.6	11.5	207372	10	AL731851	AL731851 Mouse DNA
C 39	38.6	11.5	214950	2	AC110406	AC110406 Rattus no
C 40	38.6	11.5	246528	2	AC128573	AC128573 Rattus no
C 41	38.4	11.4	66569	2	AC103837	AC103837 Homo sapi
C 42	38.4	11.4	154433	10	AL732447	AL732447 Mouse DNA
C 43	38.4	11.4	166860	10	AC115924	AC115924 Mus muscu
C 44	38.4	11.4	177967	10	AC122816	AC122816 Mus muscu
C 45	38.4	11.4	189019	2	AC129426	AC129426 Rattus no

ALIGNMENTS

RESULT 1	AC087429	88688 bp	DNA	linear	PRI 11-APR-2002
LOCUS	Homo sapiens chromosome 3 clone 2063K18 map 3p, complete sequence.				
DEFINITION	AC087429				
ACCESSION	AC087429.2	GI:20136692			
VERSION					
KEYWORDS	HTG.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 88688) Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H., F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.				
TITLE	Chromosome 3p genomic sequence				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 88688)				
AUTHORS	Liu, B., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, Y., Li, J., Li, S., Li, T., Liu, Y., Liu, N., Liu, Y., Li, W., Li, W., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-JAN-2001) 1. Chinese Human Genome Center at Shanghai				
REFERENCE	2. Institute of Genetics, Chinese Academy of Sciences, P.R.China				
AUTHORS	3 (bases 1 to 88688)				
	Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, Y., Li, W., Li, W., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Luo, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, W., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.				
	Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,				

TITLE
JOURNAL
Zhang, X., Zhang, X., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
Yu, J., and Yang, H.
Direct Submission
Submitted (11-APR-2002) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing,
100101, P.R.China
On Apr 11, 2002 this sequence version replaced gi:12039240.

COMMENT

-----Genome Center
Center: Beijing Center
Website: <http://hgclgtp.ac.cn>
<http://www.genomics.org.cn>
Contact: hgclgtp.ac.cn
-----Project Information
Center project name: 1# project
Center clone name: 2063K18
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 5% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 89950 bases at least Q40
Consensus quality: 90370 bases at least Q30
Consensus quality: 90443 bases at least Q20
Insert size: 88688; sum-of-contigs
Quality coverage: 7.47x in Q20 bases; sum-of-contigs

FEATURES

Location/Qualifiers
1..88688
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="2063K18"

ORIGIN

Query Match 12.7%; Score 42.8; DB 9; Length 88688;
Best Local Similarity 53.6%; Pred. No. 1.3;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATTATATTTTCTAGATAGCCGGTCTTTTACTACACTCAAAATAGATGAACAGATG 60

DB 1637 ATTAACTTAAACAATTAGTACCTTTATATCTTTTAAACAAGCACTTTAGAGAACT 1578

QY 61 AATGGCTGTGATCTTTTATAAGAGCTAATAAGATCATATCATTTTGGGCA 120

DB 1577 TAAGAGTAAGTCCATCTATATATAGATTAAGAACTATATATAGATTAAGCA 1518

QY 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAGTGAATAA 166

DB 1517 TAAGAACACTAATACTAGAAATAATTAATAAAATGTACAA 1472

RESULT 2
AC023353/c 187003 bp DNA linear HTG 17-AUG-2000

LOCUS Homo sapiens chromosome 20 clone RP11-775C23, WORKING DRAFT

DEFINITION AC023353 SEQUENCE, 28 unrounded pieces.

AC023353 GI:9211518

VERSION HTG; HTGS PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 187003)

Waterston, R.H.

Unpublished

The sequence of Homo sapiens clone

JOURNAL
Submitted (11-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 15, 2000 this sequence version replaced gi:7235346.

COMMENT

-----Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
-----Project Information -----
Center project name: H_NH0775C23
-----Summary Statistics -----
Sequencing vector: M13, 57%
Sequencing vector: plasmid; 43%
Chemistry: Dye-terminator Big Dye; 43% of reads
Chemistry: Dye-terminator Big Dye; 43% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 172430 bases at least Q40
Consensus quality: 176603 bases at least Q30
Consensus quality: 178873 bases at least Q20
Insert size: 18900; agarose-fp
Insert size: 184303; sum-of-contigs
Quality coverage: 3.50 in Q20 bases; agarose-fp
Quality coverage: 3.66 in Q20 bases; sum-of-contigs

-----NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1408: contig of 1408 bp in length
* 1409 1508: gap of unknown length
* 1509 3342: contig of 1834 bp in length
* 3343 3442: gap of unknown length
* 3443 5329: contig of 1887 bp in length
* 5330 5429: gap of unknown length
* 5430 8611: contig of 3182 bp in length
* 8612 8711: gap of unknown length
* 8712 11044: contig of 2333 bp in length
* 11045 11144: gap of unknown length
* 11145 15105: contig of 3961 bp in length
* 15106 15205: gap of unknown length
* 15206 17732: contig of 2527 bp in length
* 17733 17832: gap of unknown length
* 17833 21791: contig of 3959 bp in length
* 21792 21891: gap of unknown length
* 21892 23547: contig of 1656 bp in length
* 23548 23648: gap of unknown length
* 23649 28121: contig of 4474 bp in length
* 28122 28221: gap of unknown length
* 28222 34018: contig of 5797 bp in length
* 34019 34118: gap of unknown length
* 34119 38469: contig of 4351 bp in length
* 38470 38569: gap of unknown length
* 38570 44271: contig of 5702 bp in length
* 44272 44371: gap of unknown length
* 44372 51457: contig of 7086 bp in length
* 51458 51557: gap of unknown length
* 51558 58979: contig of 7422 bp in length
* 58980 59079: gap of unknown length
* 59080 65659: contig of 6580 bp in length
* 65660 65759: gap of unknown length
* 65760 74077: contig of 8318 bp in length
* 74078 74177: gap of unknown length
* 74178 82579: contig of 8402 bp in length
* 82580 82579: gap of unknown length
* 82580 92223: contig of 9544 bp in length
* 92224 92323: gap of unknown length
* 92324 99655: contig of 7332 bp in length
* 99656 99755: gap of unknown length

	*	99756	108570:	contig of 8815 bp in length
	*	108571	108670:	gap of unknown length
	*	108671	115388:	contig of 6718 bp in length
	*	115389	115488:	gap of unknown length
	*	115489	123778:	contig of 8290 bp in length
	*	123779	123878:	gap of unknown length
	*	123879	132327:	contig of 8449 bp in length
	*	123238	132427:	gap of unknown length
	*	132428	145195:	contig of 12768 bp in length
	*	145196	145295:	gap of unknown length
	*	145296	158332:	contig of 13037 bp in length
	*	158333	158432:	gap of unknown length
	*	158433	171052:	contig of 12620 bp in length
	*	171053	171152:	gap of unknown length
	*	171153	187003:	contig of 15851 bp in length
FEATURES			Location/Qualifiers	
	source		1..187003	
		/organism="Homo sapiens"		
		/mol_type="genomic DNA"		
		/db_xref="taxon:9606"		
		/chromosome="20"		
		/clone="RP11-775C23"		
		1..1408		
	misc_feature	/note="assembly_name:Contig29"	1509..3342	
	misc_feature	/note="assembly_name:Contig30"	3443..5329	
	misc_feature	/note="assembly_name:Contig31"	5430..8611	
	misc_feature	/note="assembly_name:Contig32"	8712..11044	
	misc_feature	/note="assembly_name:Contig33"	11145..15105	
	misc_feature	/note="assembly_name:Contig34"	15206..17732	
	misc_feature	/note="assembly_name:Contig35"	17833..21791	
	misc_feature	/note="assembly_name:Contig36"	21892..23547	
	misc_feature	/note="assembly_name:Contig37"	23648..28121	
	misc_feature	/note="assembly_name:Contig38"	28222..34018	
	misc_feature	/note="assembly_name:Contig39"	34119..38469	
	misc_feature	/note="assembly_name:Contig40"	38570..44271	
	misc_feature	/note="assembly_name:Contig41"	44312..51457	
	misc_feature	/note="assembly_name:Contig42"	51588..58979	
	misc_feature	/note="assembly_name:Contig43"	59080..65659	
	misc_feature	/note="assembly_name:Contig44"	65760..74077	
	misc_feature	/note="assembly_name:Contig45"	74118..82579	
	misc_feature	/note="assembly_name:Contig46"	82680..92223	
	misc_feature	/note="assembly_name:Contig47"	92334..99655	
	misc_feature	/note="assembly_name:Contig48"	99756..108570	
	misc_feature	/note="assembly_name:Contig49"	108671..115388	
	misc_feature	/note="assembly_name:Contig50"	115489..123778	
	misc_feature	/note="assembly_name:Contig51"	123879..132327	
	misc_feature	/note="assembly_name:Contig52"	132428..145195	
	misc_feature	/note="assembly_name:Contig53"	145296..158332	

		/note="assembly_name:Contig54"
misc_feature	158433..171052	/note="assembly_name:Contig55 clone_end:T7 vector_side:right" 171153_..187003 /note="assembly_name:Contig56"
ORIGIN		
Query Match	12.7%; Score 42.8; DB 2;	length 187003;
Best Local Similarity	53.6%;	Pred.No.1.2; Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
OY	1 ATATTATTTTCATTAGATGACCGGGTTTTTCTAACAATAAGTGGAAGAATG 60	
Dd	119465 ATTAAACTTAACAATTGTAGTAGCCTTTTATATCTCTTTAACCAACGAACTTATGAGAACT 119406	
OY	61 AATGGGTAGGACGC GTTATPAAAAGATAAAGTAAAGTACTATCATCATTTGGCCAA 120	
Dd	119405 TAAGGTATGACCATCTATTAATPAAGAAATGAAACACTATATATPATAAGATTAAGCAA 119346	
OY	121 TAAGGAGAGGAGAGATTCAGCAAAACAGTGTGCTTCAAGTGAAAA 166	
Dd	119345 TAAAGACACTAAATCTTAGAAATTAATTATAAATAATGTACAAA 119300	
RESULT 3		
AEO14841		
LOCUS	AEO14841	252420 bp DNA linear INV 07-OCT-2002
DEFINITION	Plasmodium falciparum 3D7 chromosome 11 section 6 of 8 of the complete sequence.	
ACCESSION	AEO14841	AEO14186
VERSION	AEO14841.1	GI:23496259
KEYWORDS	.	
SOURCE	Plasmodium falciparum 3D7	
ORGANISM	Plasmodium falciparum 3D7	
REFERENCE	Buxaryota, Alveolata; Apicomplexa; Haemosporidii; Plasmodium. 1 (bases 1 to 252420)	
AUTHORS	Gardner,M.J., Hall,N., Fung,E., White,O., Berriaman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.F., Bowman,S., Paulsen,I.T., James,K., Eilen,J.A., Rutherford,K., Salzberg,S.L., Craig,A., Kyes,S., Chan,W.-S., Nene,V., Shalim,S.U., Sun,B., Peterson,J., Angiloul,S., Perlea,M., Allen,J., Selengut,J., Hatt,D., Maier,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H., Frunholz,L.M., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Batteilli,B.	
TITLE	Genome sequence of the human malaria parasite Plasmodium falciparum	
JOURNAL PUBMED	Nature 419 (6906), 498-511 (2002) 12368864	
REFERENCE	2 (bases 1 to 252420)	
AUTHORS	Gardner,M.J.	
TITLE	Direct Submission	
JOURNALS	Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA	
FEATURES	location/Qualifiers	
Source	1..252420	
	/organism="Plasmodium falciparum 3D7"	
	/mol_type="genomic DNA"	
	/isolate="3D7"	
	/db_xref="taxon:36329"	
	/chromosome="11"	
	complement(85..120)	
	/note="AT-rich"	
	/rpt_type=tandem	
	/rpt_unit="(ta)n"	
	224..253	
	/rpt_type=tandem	
	/rpt_unit="(ca)n"	
	360..399	
	/rpt_type=tandem	

Gene	Repeat Region	Repeat	Repeat Region
gene	complement(360..399)		
	/rpt_type=tandem		
	/rpt_unit="(ta)n"		
	<419..>3287		
	/locus_tag="PF11_0334"		
mrna	join(<419..3121,3279..>3287)		
	/locus_tag="PF11_0334"		
	join(419..3121,3279..3287)		
cds	/locus_tag="PF11_0334"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="AA035918.1"		
	/db_xref="GI:23496260"		
	/translation="MGTPDNTDEYNTITLHLERRRKRNLQDVSHNKDPEDLKLYLYNACLFASNTDIYQVEYINRLSINSYNNCDYDQIKSCYLSNDPFLKYNNYNDPFLVLENDNTDSSSYHLNEMKELEELSEFVYIHKMKIKKEFVNDISRRRSIGCTGTVAMSKERKTENNLSLSISYNNNNNNLSTNPOKONSHLFTLLERKNNKKNKNNLYLSHIGDKNEBSIYNHEHTSYNYLPYTPPMKKNNIPSTCQNSRHSISITHMDPFLKCDISRMRLKNSGNSYFCMLQNIIEKQNDIYAMRYKYLDDVYKKSRNATYDNYLSGMDLNTYKKKKKKKKIKKDKSKSKKKNEMYDKNIYNNNTNLIANTSNSHYEYELISNNIINIDEANTLENYIDMEIYNNNLHTSDNNKKDITKINBPYLVQINKKKNIIRKLSSVENDIYIKTMINENKIKSNQORIDRRDKNALYQMFSCIGASISIVYEDLPRLGRKRPFYLLTCVILIPCSYIPIGPILOIELISGOLSGOGLNLSFLAKKIKGIGLSFILLFLFKRNNNSYITTYLVIVANSYWRPLPMDIKCEKINKQCTCVSDIKRKYEMKEKODITKPKYDNIQYKDLDSQNGLDKHPYRNLSTHNSNDKLQDISHSEKKNHNDLNIYMSKRRPEPVNNNSCVSISTMEIYILPEFKYISPMWIKLSLFWILLVFLKTIETMSLNQSHIYLSISFLIIVLQILILYLYKLEPEKYTONEMAKIKYNTCLFFSKELYLHNFLPIIKITICLVLFVNSGSGINIVPSSYTNIGDNIKQSYIILGGSFIVITTYLTYYLICIHLHNNNKNNINNYDDIYLIANYFLIDYNNLPINAHISY"		
	complement(677..706)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	957..976		
repeat_region	/rpt_type=tandem		
	/rpt_unit="(taa)n"		
	1032..1068		
repeat_region	/rpt_type=tandem		
	/rpt_unit="(taaaaa)n"		
	1453..1478		
repeat_region	/rpt_type=tandem		
	/rpt_unit="(a)n"		
	complement(1520..1556)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(1584..1612)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(1680..1717)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(1737..1766)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(2814..2846)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(2936..3100)		
repeat_region	/rpt_type=tandem		
	/rpt_unit="(cataa)n"		
	3015..3046		
repeat_region	/rpt_type=tandem		
	/rpt_unit="(caa)n"		
	complement(3129..3184)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(3356..3496)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	complement(3558..3651)		
repeat_region	/note="AT_rich"		
	/rpt_type=tandem		
	<3785..>4738		
gene	/locus_tag="PF11_0335"		

mRNA	<3785. .>4738	
CDs	/locus_tag="PF11_0335"	
	3785. .4738	
	/locus_tag="PF11_0335"	
	/codon_start=1	
	/product="hypothetical protein"	
	/protein_id="AAN35919.1"	
	/db_xref="GI:23496261"	
	/translation="MNILKRLINNI FYGKYRKYDTOTNKENNINIKNKIYKCHILQCL TMMQFLFKYFTTTFIPILISLHVEENKLYHKLKONNYLSYKSGOYLIHNIQ YIYKHINDEFNIRKVDHDKNTEKKKSIOKNNNIOILNNTLHEFKPKGYGQDDN XIKHNIWNTLLIYITLPTIIFILPSFSFYFGKPKFNFINSPKLNININIDQKI RKVEVFLTRSPFFFEELLPVQYIMPSTIWHINKLTKTSFIQSLKGEYENCENIID KKFLPYNAQSFESYILGIGCGSLEHFTOKKKKRIIDKOE"	
repeat_region	/note="AT rich"	
	/rpt_type=tandem	
	complement(4077. .4130)	
repeat_region	/note="AT rich"	
	/rpt_type=tandem	
	complement(4119. .4217)	
repeat_region	/note="AT rich"	
	/rpt_type=tandem	
	complement(4296. .4333)	
repeat_region	/note="AT rich"	
	/rpt_type=tandem	
	4720. .4757	
repeat_region	/rpt_type=tandem	
	/rpt_unit="(cataa)n"	
repeat_region	complement(4750. .4846)	
	/rpt_type=tandem	
repeat_region	/rpt_unit="(ca)n"	
	4762. .4794	
repeat_region	/rpt_type=tandem	
	complement(4874. .4917)	
	/note="AT rich"	
	/rpt_type=tandem	
	complement(4910. .>6244)	
gene	/locus_tag="PF11_0336"	
mRNA	complement(join(4910. .5940,6115. .>6244))	
	/locus_tag="PF11_0336"	
CDs	complement(join(4910. .5940,6115. .6244))	
	/locus_tag="PF11_0336"	
	/codon_start=1	
	/product="hypothetical protein, conserved"	
	/protein_id="AAN35920.1"	
	/db_xref="GI:23496262"	
	/translation="MANTRTGSALIKNGSNFOYIISNIRSKSYDSPYWKEKCFALTS BSIIQAINLVKVGDTGGNRKPTRFLCLILKLQIQDDKDIYEVINKDFPYRLAL GIIFYRLIKGSLSEVNLHLEPIFYRKIPMRLONGTEKILYMDVFNCLILNLFVRL DPTFLTKVOLENNLLQRKILLEYKALNTISNLEPKNNEEDTHNNKNKINNGKNK YQRMILRRHKIKYNSSEDDHNYVYHKEGVYDKRKKGNKIYHKKYNNKQDDERHHH SYSDMDMRKSRKKKKRTHSKTDHKKKKKKRGRYSSPSKISFTSSEYSVQKKKEHHK KKKKLFLFKQKKKKYHSSSNHSDYDNKKDELIDKMSIRKLQMKPLK"	
repeat_region	complement(5025. .5094)	
	/rpt_type=tandem	
	/rpt_unit="(a)n"	
repeat_region	complement(5145. .5218)	
	/rpt_type=tandem	
	/rpt_unit="(a)n"	
repeat_region	complement(5368. .5388)	
	/note="AT rich"	
	/rpt_type=tandem	
repeat_region	5945. .5983	
	/rpt_type=tandem	
	/rpt_unit="(a)n"	
repeat_region	5985. .6016	
	/rpt_type=tandem	
	/rpt_unit="(cataa)n"	
repeat_region	complement(6046. .6113)	
	/note="AT rich"	
	/rpt_type=tandem	
repeat_region	complement(6251. .6290)	

[illegible]

This is a complete sequence.

Genes were identified by a combination of several methods: Gene

Game

FEATURES

source

genescan (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHM (<http://genemark.biology.gatech.edu/Genemark/>) trnascan-se (Sean Eddy, <http://genome.wustl.edu/eddy/trnascan-se/>), searches of the complete sequence against NCBI nonredundant protein database (<http://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.

Location/Qualifiers

1..39855

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/sub_species="japonica"

/db_xref="taxon:39947"

/chromosome="4"

/clone_1ib="CUG1-OSUNBA"

/clone_1ib="CUG1-OSUNBA"

complement (3866..6899)

/gene="OSUNBA0064D20.1"

complement (join(3866..5543,5638..6485,6573..6899))

/gene="OSUNBA0064D20.1"

/codon_start=1

/protein_id="CAE04217.1"

/db_xref="GI:32489561"

/db_xref="UniProt/TREMBL:Q7X884"

/translation="MARSGAPLDGSLVPPSRIVSERQAGLPFRFMPESATIGREIVSLG
ECGPARYPGSVFLPFSMAGLVPPSSFPMDVLEFYDLQMAHLTPNAVMTALFTH
LGMFTGPRGVLNRYTPCVLRKMDKMSDPYTIADDEARLRISQPPASQPMW
PVLIDGPRVLDRLAGLRFQSLTGAMVGVILHRIAPLQORAGAMPTYSSEPMW
THQVRMDWAPEDFKIVORVNLSSVESLTPQVLPLCSPPDASITITLQAVAS
ERRAPGHGAGSHRGEQSTPGQGSAPRGQPSGPARARKRGKEGPPSP
RCGGAAPASRSBEGATPTSGEGERKKRLRMGTGSPRNLISPPRSFNNRPS
AIEPSRSHSKSGSEADPATTEARRESDRREADRLREABEARAROTEEA
AREEGALROAATTSSEATRETDROBARLREABEARAOVAROAEBAAREAR
QAEATTSATDEAGASLGPPTCDADKRPEDIPSSGISGSPSAASSRPL
PVPAPLSAEPLQALVANTVLDGSLAQVALQAEBAELDAMARVEGRSVEA
MVEVGRKARRHVSLEARKYGLAKEVEERGAALJATVMBADPTLQVGSW
BAELGKLDAAAGVLDAAAREERAAETEVASRRERALEARMALEERACAVEEDLA
DEBAVAVIREATLTHEACAEBESALRLBEDALTERERALEABEAALQRLADSLR
BAAREBOACRLNLEGARERAAINORAELEAREERLIDARAGGAAGESDITALLAA
AHTTIDLOCALDSSAGVEBALRAGVEPGHFRDVSULDAGROAGLGMGRTRYKTA
ANOGGLAQRLSETAGTLQRLPSELEKTISSSRDLA"

complement (10918..12816)

/gene="OSUNBA0064D20.2"

complement (join(10918..11100,11189..11344,11862..11979,
12107..12816))

/gene="OSUNBA0064D20.2"

/codon_start=1

/protein_id="CAE04218.2"

/db_xref="GI:38346571"

/db_xref="UniProt/TREMBL:Q7X902"

/translation="MMGFSCSLIDLQGMILHMPRPGRGAVOTPPSPATRLSATL
PDLATLTLSSLRGMOVOTSVLSKRRHLMRSVPELDDQREPAASNNATISSDL
EKEDPADVNLARCGSPAKLDLFRILCDRHSLSSLDTPDMIRGKLCSREPHLH
FDYRDSYLLMEHKLGSNYGCLTKLTLTVSLHECMEMHITVCTLLLEVLNRSLEY
LGEITHPKLKNLHGPVAMPSPVIRASIRFSVFPCKCNFLVLFNVTLSLSEGRFM
VLEFNOVEFHAFKNTLRLDRCSRNNELRLHLNOSPMLKLTVCCKFSGSL
EMKSSQHKNOVNCRLKSTELIYKXIDIVRELVDLLDVSGLHPNTALTKI"

complement (22800..27420)

/gene="OSUNBA0064D20.3"

complement (join(23800..24075,24708..24800,24919..24998,
25802..25967,26575..27420))

/gene="OSUNBA0064D20.3"

/codon_start=1

/protein_id="CAE04219.2"

/db_xref="GI:38346572"

/db_xref="UniProt/TREMBL:Q7X801"

/translation="MEPVAKRARTGGGSAVADRLSLSPDLALHAWMSLPAQWVO
TCVLAKRWHLMSVPSNLDSREFLPIYDRQKEDPTPLALMFHAPLTDASIR
ADYAVGKHGRHVRWTRCGIKYCPBVLIDVAVTVDSRYRLPDLASGSCGLHLSYV
AAHSGARQVRDSCPYLRCLERCLTKSHHESSTLANLVIEDSGSDSLAIEPR
LMSLRVLEFHTYKNGSLNGANSVLEASVAKSRTSPGSAAMLGCLFVSSLEL
KGLQELVFLVGLQIKQSWRLATTLVLESEFOALIEHFKSQSFHNLRTLSNCF
QAEADVDAAADSMQVRPIISLQRLKEDKVFVQBPDELSPFNKRAKCCSIASSA
NVGKAMEEHEHNDEHGRRRPSLAAPSQSPAAGLAAPPIVSTRGRPLHGEARQR

gene

CDS

CRRLSIFAGLSPPHVPAGRRRHIGDVIQSDAVAQ"

complement (29612..32646)

/gene="OSUNBA0064D20.4"

complement (join(29612..29827,29918..30088,30550..30590,
31036..31096,31762..32646))

/gene="OSUNBA0064D20.4"

/codon_start=1

/protein_id="CAE04220.2"

/db_xref="GI:38346573"

/db_xref="UniProt/TREMBL:Q7X877"

/translation="MDRLSALPDGLHAWNSFLPARONAQTCVLSKRWHLMSVPSL
NLDIRFENLSKEDEDDDETGMKQFIVANLHMFHAPLTDFTVCGAIVOAVGN
NSRRHNRNDVDVPAVDNRSDVGRMTIRGGIKYCPRYLDVAVAPSGSGCPPDGS
GSCRPRRLRLSCCALDSGRARQVRDSCPYLRCLERCLTKSHHESSTLANLVIE
CCTGCSLSLISAPRLASLCLDLSYGARKNVSILNTLSLVASSTLVNFQISPSGEM
LLCGLFVNTNLELAHAKLGVLAGAKLSAVGSAKGRERERKYGCCVANKVYLHE
KPDFRSSFNNTLRLSLDCCFQGMGLKEKFPALRLKCCNLEKLTLOHGMFLSGST
QRAEMETRTSGLQSDQOVTFHCQKLTLEIKHDSLELCCDHLFQLMMGFMKDLKKA
TILITKI"

complement (37808..41919)

/gene="OSUNBA0064D20.5"

complement (join(37808..37912,38701..38871,38985..39135,
41339..41919))

/gene="OSUNBA0064D20.5"

/codon_start=1

/protein_id="CAE04221.2"

/db_xref="GI:38346574"

/db_xref="UniProt/TREMBL:Q7X6F3"

/translation="MGPRRRAPAVLEINVTBSPYBGHYDGYTMPDLITSRAMR
LTRLISLHAMLDGQGLDGCPLBEDLRACAMAGFRRIRGSLRTIYLHYTG
GGGDGADDEEBEETLVISAPRLASVVKITYSACRGVSGDSTAGSLVASIRVG
RRRRRLPTVEENTLACGMVNTLSSKYVRQDGTPLVSGVNSVASISDVSNOM
SPSGEAMLIGLFCSTNILEKGTAMALIDBELDFPFPNNLRNLSLHCCIRKDNLS
DRFVALGLQKSLNLEKTLQDFEASWCNNTSVLGIQVLLFLFPIILHVASDL
N"

complement (45448..46707)

/gene="OSUNBA0064D20.6"

complement (join(45448..45516,45762..46130,46270..46707))

/gene="OSUNBA0064D20.6"

/codon_start=1

/protein_id="CAE04222.2"

/db_xref="GI:38346575"

/db_xref="UniProt/TREMBL:Q7XSX6"

/translation="MADRTTATMTPAASSRRKARVPTTGAGGGGGEGRIEGLPD
ELLISLISCTTRQAVQTSVLSRRRHLMRSTPRPDVLAIFARPPESAFLHLHGRG
STDEWERLGRFTARLMTYHAFLVDAELRVAATPPHRADEVSMARFGLDLSGCPAL
VDMELDRCKCFHELSATLRLSAMESCLMMKRPSTGTDIVSVVAPRLAYLRLLTF
GHGCKVRFPSGDSISSEVSIRGFNLINLRLRLMMENVTTLRLSGFGPTSVSARMF
RKFPSPGSDNLNAG"

47424..50668

/gene="OSUNBA0064D20.7"

join(47424..47697,48191..48465,48691..48855,48923..49012,
50294..50668)

/gene="OSUNBA0064D20.7"

/codon_start=1

/protein_id="CAE04223.2"

/db_xref="GI:38346576"

/db_xref="UniProt/TREMBL:Q7X8C0"

/translation="WGGGGGGRNVHNDHDDDDWELAGALADTVLVLGVKVGSGKS
ATANSILGDEAFESKCSYAGVTOCTCKSTVQDCLRTINVIDTPGLPMDIYAD
VREIVKGDMAKDGILHAMLVFASATSSFCDECTIETLSFEDKLLDHMLIVETR
GDEVGSTSGKMLSDSAPTYLDITLKPENRVYLFKKTSSTTOROKRMILDAVD
FVVSNHGKPPSNQIFTOIQEYHROKANSSEVYSMOTBSYISLITKRYBEKLNGT
ILRMQOLKQEARLDIONETTKALISSEEDIRLRSLSEKVAQESNNANENRFRF
SEKASKQEKQTEAEIOKLKEMKEDKEEBEEIRRLRDLLEKREBROKOSGCIIL
"

complement (51830..56465)

/gene="OSUNBA0064D20.8"

complement (join(51830..52739,52843..52970,54357..54818,
56012..56229,56348..56465))

/gene="OSUNBA0064D20.8"

/codon_start=1

/protein_id="CAE04224.2"

```
/db_xref="GI:38346577"
/db_xref="UniProt/TREMBL:Q7X729"

Query Match      12.3%; Score 41.4; DB 8; Length 9985;
Best Local Similarity 51.3%; Pred. No. 2.8;
Matches 96; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 37 AACCTAAATAGATGACAGATGATGGTGTGTGACCTGTTATTAAGAGAGTAA 96
DB 83615 AATTATGTAACCAACAACTACTGCTGCTTTCTTCAATTAAGATTAATGTGTGT 83556
QY 97 AATACTATCATCATTTTGAAGCAATAGGAGGAGAGATTGCAACAGTGTGCTAC 156
DB 83555 ATTAATATCTGATTTGTGATTAATTTAAGATTAATTAACGATGATTAAGTCTTTA 83496
QY 157 AAGTGAACAAGATTAACTAAAGTACCCCTCTTCAAGATCAATGCCACAGTT 216
DB 83495 AAGTGAACAAGTTGTTCTACAAAGATATATCTCCCTCATTCCTTAATGTTTGACACCGTT 83436
QY 217 GAGCTTT 223
DB 83435 GACTTTT 83429

RESULT 6
OSJN00292/c 121130 bp DNA linear PLN 10-FEB-2004
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: P0076017,
DEFINITION complete sequence.
ACCESSION BX548156
VERSION BX548156.1 GI:32479667
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1 Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q., Hu,X.,
Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J., Zhang,L.,
Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y., Li,C.,
Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K., Zhou,B.,
Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P., Pu,G.,
Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F., Tu,Y.F.,
Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H., Chen,X.Y.,
Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W., Wang,L.J.,
Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L., Zhu,F.H. and
Hong,G.F.
Direct Submision
Submitted (08-JUL-2003) Han Bin, National Center for Gene Research,
Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn
Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: P0076017.
Web site: http://www.ncgr.ac.cn
----- Summary Statistics
Assembly program: phrap
-----
This is a complete sequence.
Genes were identified by a combination of several methods: Gene
Prediction programs including Fgenesh (http://www.softberry.com/),
genscan (http://CCR-081.mt.edu/GENSCAN.html), GenemarkHM
(http://genemark.biology.gatech.edu/Genemark/), tRNAscan-SE (Sean
Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
complete sequence against NCBI non-redundant protein database (nr)
(http://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
FEATURES
source
1..121130
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"

/db_xref="taxon:39947"
/chromosome="4"
/clone="P0076017"
/clone_1lb="RGP-06JNP"
/complement(14269..7724)
/gene="P0076017.1"
complement(join(4269..4451,4540..4695,5213..5330,
5458..6125,7695..7724))
/gene="P0076017.1"
/codon_start=1
/protein_id="CAE02503.1"
/db_xref="UniProt/TREMBL:Q7X902"
/translation="MASTFRKTKTGMILHPRPPRGCAVQTSPSPATDRLSALPDAL
LHTIILSLKRGQVOTSVLSKRWRLHMRVSPCLDIDREFPAASENMAISRDLKFE
FDADVLAVYRCGSPAKLDITFLRLICDRYHLSLSDTDWRIRGLKCSPREPHLPDYR
YDSYLLNEMHKGSNYGLTKLITVNSLHCEMBEHTTCTTLEVLNRCGLSYLOEI
THPKLKNLVGHGPAVMSLVRSIRFSVFPCKNPLNVLFNVTLSLELSGPREMVLLE
FNQVFNPAFKNIRTLILDRCRPSYNNELRHLLQNSPNIKEKTYVHCCKRSKSLERK
SQHKQVNCRKLSKSTETITVKDIDVRELVDLILVSGHLPKNTALTATXI"
11169..13339
/gene="P0076017.2"
join(11169..11467,12805..13339)
/codon_start=1
/protein_id="CAE02504.1"
/db_xref="UniProt/TREMBL:Q7X801"
/translation="MAPIPSKLSQGRPIAPVWSHLYARKNPPOESVLICGTVPH
RHNPIASFHNPFYFOTHRHNSGFGSGNSPPKSTNNVCIGICIGIAPORARIPRI
GMSYTBEGRRRAATTEFGRRORRAYGEORRSGRQLATGLTALLPRAKERGSEEG
EKEMHRAVETVTSGRRRGSELGIGRKGKETAAPVHQIATTPASPAVAVEGG
EDGSGKAAKARWNRNBERBDGARFNRQCFRKEPTLAVKESBLAVIDACOSODG
R"
complement(16366..20771)
/gene="P0076017.3"
complement(join(16366..16494,18059..18151,18270..18367,
18906..18969,19067..19237,19926..20771))
/gene="P0076017.3"
/codon_start=1
/protein_id="CAE02505.1"
/db_xref="UniProt/TREMBL:Q7X877"
/translation="MEPAKRAKRTGGGGSAVDRLSLPDLAHAVKSSLPARQMSR
TCVLSKRWHLHMRVSPSLNDSREFLPIYDMQKQEDFTTLMFNHAPTDAFIR
ADVAAGKGRHVDWIRCGIKYCPVLIVAVTVSRAYLPIAAGSGCRPPIDGS
ALHSGPAROVDRSCVRLCRLCTKFSHSESTLNRLVIEDSGSDSLAISAPR
IASLPIVAFLEHTYKNGVSLNAGNSIVASVAVSGRTSPBEAMILCGLPVSSELE
KIQRLATLHKEFKDSQSPFNIRTLISLNCFOABGDLKDFRALGKLEMCNLEKUT
LOHCWFLIGSTQIRAKRTKRTSGLCLENSESETGADVDVAALDSQVPRPITLSJORLED
KVTFOEBDELSMFNLRACCSIASANVVISRLKRGDEGETDLDSDVPDPHENT
SQCPLKATYIEIKYKLI"
complement(22963..25997)
/gene="P0076017.4"
complement(join(22963..23178,23269..23439,23901..23941,
24387..24447,25113..25997))
/gene="P0076017.4"
/codon_start=1
/protein_id="CAE02506.1"
/db_xref="UniProt/TREMBL:Q7X6F3"
/translation="MDRLSALPEGLHAWMSFLPARONAQTCVLSKRWHLHMRVSPSL
NLDIDREFNLSKEEDDDDETKRKXKDELIANLMPHNAFTLDTFTVTCGAAVQAAVGN
RSRRHRRANDVDPVAVDNRSDVWRIRGGIKYCPRYLDVAVASGSGCRPPIDGS
GSCFRLRLRLRSCALSDSGPAROVDRSCVRLCRLCTKFSHSESTLNRLVIEDSGSDSLAISAPR
GICGSLSLAISAPLALSLCIDLSGAYKNGVSLNTIIVASVYTNLNPISPBEBAM
LICGLPNTNILELHIAKLGVDKAKISAVSAGSGRERKPTGCCCNKVKTLHIE
KPFKSSFNNAKRTLSLDDCFQMGDLKKEFLALGKLEKCPLEKUTLOHCWFLIGST
QRAEMERTSGLCQSDQDQVTFHCKLILTKHDSLECDHQLPQLMWGFWKDKKA
TIIITKXI"
complement(31159..35270)
/gene="P0076017.5"
complement(join(31159..31263,32052..32222,32336..32486,
```

```

34690..352701)
/gene="P0076017.5"
/codon_start=1
/protein_id="CAE02507.1"
/db_xref="GI:32479672"
/translation="MGPRRRAPATVLEINVTTPSYPGHYDGYTMPDLIITSRAMR
LTRHLSHAMLDGGEGOLGDCPLLEDIALRCMAGDPRIKSGSLRTIYHYTGC
GGDAGGDEEBEERTVLSAPRLASVRKITSYCRHVSFSDSTADSLIVRSG
RRRRRAPPTGVBAVLAMVNVTTLSKVTYRGQGTSLVGNVSLVSRASIDVSNOM
SPSGEAMLLGALFSAITNLEKGTAMALDELDLDFPLPNNIRNLSLHCCLDKNLS
DRFKALGRLLQKSLNLEKTLTDFFWASWCNRTSVLGKIQVLLVLLPLPILHVASDL
N"
complement(38799..40058)
/gene="P0076017.6"
complement(join(38799..38867,39113..39481,39621..40058))
/gene="P0076017.6"
/codon_start=1
/protein_id="CAE02508.1"
/db_xref="GI:32479673"
/translation="MADRTTTHMIPAASSSRKRRVPTTGAGDGGGGEGRLGSLPD
ELLJLISCLITRQAVQTSVLSRRKRMHMRSTPRDVLAEFARPPSSAPMLHGRG
STDPEWRLRGFTARLLMTHAAVLDAFLRVATPPHRRADVSMGSAFAGDLISGCPAL
VDMELDRCKCFHEHLSATLRSLAMESCLMNRPGTGMDRTVSVAADRLVLRLLTE
GHGDKVFRFESGDSISEVYIRGFTNLINLFRLLMMPVTTLRISGFGPVSARMF
RKEPRPDSNIALG"
40499..44019
/gene="P0076017.7"
join(40499..41048,41542..41816,42042..42206,42274..42363,
43645..44019)
/gene="P0076017.7"
/codon_start=1
/protein_id="CAE02509.1"
/db_xref="GI:32479674"
/translation="MEEVYMLDLEIMYTYKLAARTGSAHNAHLFVSDTSSFR
GKLSAACINHYLSPSSSLIPTRKFPVSATPVISQSLSDPIMGGGGGRNV
DNHDDDDMELAGALADVLIVLVGVKSGSKATANSITIGDEPESKCSAGVOT
COKKSTVODGLIRIVIDTPGLFMDIKAEVDYRRELYKCMDAKACITAMLVFS
ATSRFSCDEKLETLSKFPGLIDHMLIVTRDDEVGESTMKNLSDSAPYLDH
ILKLFENVRLEFNKTSSTQDRQAKRMKLDVADVSVSNHKKPSPNSOLFQIQVH
ROKDNSEVYSMOETDSYISLITTKVEKLNGLILMEQCLKEQEARLDIONMTK
ALRSEBIDIRRLRSLSEKAKOESNNARENKFRSEKSKSQEQETAEIQLEKEM
EKDREEREIRLDDLEKEBERQKQSGCIIIL"
complement(45181..49816)
/gene="P0076017.8"
complement(join(45181..46090,46194..46321,47708..48169,
48735..48764,49363..49580,49699..49816))
/gene="P0076017.8"
/codon_start=1
/protein_id="CAE02510.1"
/db_xref="GI:32479675"
/translation="MDVESCASSPPDDAADMGRGPCPERRHGGRAAVFVLGIQAF
EIMATIAVGNMLITYVGEHMPPLSOAVNVFVGTIPLSLGPFSDSYLGFWT
MIFGFVELSVFNDPTATGFTILISVOAHLPOLKPPCMNMAATGGCGOARCIASL
PFAALYIVLGSCKKNNIAGADOPFAAAGGGAANAADKRTISYNSYFSPCA
GELVATLAVVQTHSGMDVGFISAAAGLVSLVSGAAYRNKPEGSIFPTIAR
VFAAAYTRKQICPSSSDPNAVAGCEPHLAGSGFRASKRFLDKACIRAGAGPN
TRPSPFMRICTAAYROAKTLVAVIFACTIVFVTLAQOTFVQOQSANDTLCG
AGSPRIPIPASLOAIPLYMLALVPAVELLVLIMBRATGASGTPLQRIQVGLCTV
PLSMVAATVEHRRRDLISAGGAPRAVSUIMVOPILVPSVEMFATGGLIEFPYK
QARGAMQSFTRNLTYCSYAFGYLSSVTVSVNRYTASRGCGGCGHGLDNDLIDKR
LDLFTWMLAVLSINFCYLLCARWNSGADGCDASASAVAAVAGDNGKEII"
complement(64136..69177)
/gene="P0076017.9"
complement(join(64136..64335,64429..64596,64680..64899,
65493..65683,65791..65884,66179..66422,67236..67333,
67431..67522,67617..67731,67812..67958,68953..69177))
/gene="P0076017.9"
/codon_start=1
/protein_id="CAE02511.1"

```

```

/db_xref="GI:32479676"
/translation="MSQRAVAARAAAGGQVQGAARQAGGGLGQITAGIYNAV
WYPAKFPGRKPPAEFGKMLNSLFOKEEPLDMNMYLSENKFNDSMEDLVNHEAN
IPVAVWGPTSTRHTLTYLPSBAIKNNSLAHYFAASGVVPVPTDEYQSSAFGR
THPVVAYLRKPKDCKKSLDQSSNSQPPKKNKQSDVDQGPVEIYSWKPVVTTI
NLVDDFTNNTDLNVDSNNYYPTVFNFEFWLIDKLIALNEVEELPLNEVGPISM

Query Match      12 3% Score 41.4; DB 8; Length 121130;
Best Local Similarity 51.3%; Pred.No.2.8;
Matches 96; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

37 AACTCAATTAAGATAACAGATGATGATGCTGTAGTGTATTAAGAGATTAATA 96
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAATTAGTACCAACAACATGATGATCTTTCTTATATGATATGATGTGTGTGT 76907
97 AGATACATATCATTCATTGAGGCAATTAAGGAGGAGAGATTCAGCAACAGTGCTTAC 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
ATAATATATCTGAATTTGTGATTAATATTAAGAAATGTAACGAATGATTAAGTCTTTTA 76847
157 AAGCGGAAAACAAGTAACTTAAGTACCCCCCTCTTACAATGATCAATGACCACTT 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AAGTGAAGAGTTGTTCTACAAAGATATATCTCCATTCCTTAATGTTTGAACACCGTT 76787
217 GAGCTTT 223
||| |||
76786 GACTTTT 76780

```

```

RESULT 7
AL928605
LOCUS
DEFINITION
ACCESSION
AL928605
VERSION
AL928605.22
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209024)
Direct Submission
Submitted (08-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humgvery@sanger.ac.uk
On Sep 9, 2004 this sequence version replaced gi:50949682.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a paired quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgvery@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the rare assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

QY 257 GATTAAGCCTGTTCAATAATTCCTTTGCAAGCTCTGACTTACCAAGATGCTGCTAC 316
DB 114604 GGTACCTTTTAAATTCATTTCCCAAGCTTACTTCTTAAAAAGAGTCAGTTT 114663
QY 317 AGACAGAAAGATGGCATT 334
DB 114664 AGAAAGAAAAAGGTATT 114681

RESULT 10
AC026155 174294 bp DNA linear HTG 26-JUN-2001
AC026155/c Homo sapiens chromosome 12 clone RP11-180F6, WORKING DRAFT
LOCUS
DEFINITION
SEQUENCE, 12 unordered pieces.
AC026155
AC026155.23 GI:14547355
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 174294)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Albrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbier,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowen,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Daborn,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Diepert,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falle,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homai,F., Howard,S., Huber,J., Hulvik,S., Hume,J.,
Jouchon,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korva,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W.,
Lousesed,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokemko,S.,
Ogih,M., Okunou,G., Orsanye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,K., Shoshari,N.,
Sisson,I., Sodergren,B., Sonaite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmari,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wisczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 174294)
Worley,K.C.
Direct Submission
Submitted (21-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:13811703.
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: HALP
Center clone name: RP11-180F6
Sequencing vector: pBluescript
Sequencing vector: M13, L08821
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 176003 bases at least Q40
Consensus quality: 179810 bases at least Q30
Consensus quality: 181603 bases at least Q20
Estimated insert size: 178132, sum-of-contigs estimation
Quality coverage: 0x in Q20 bases, agarose-fp estimation
Quality coverage: 8.6x in Q20 bases, sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 35024: contig of 35024 bp in length
* 35025 35124: gap of unknown length
* 35125 56161: contig of 21037 bp in length
* 56162 56261: gap of unknown length
* 56262 77352: contig of 21091 bp in length
* 77353 77452: gap of unknown length
* 77453 103958: contig of 26506 bp in length
* 103959 104058: gap of unknown length
* 104059 119954: contig of 15896 bp in length
* 119955 120054: gap of unknown length
* 120055 130550: contig of 10496 bp in length
* 130551 130650: gap of unknown length
* 130651 142704: contig of 12053 bp in length
* 142704 142803: gap of unknown length
* 142803 149662: contig of 6759 bp in length
* 149663 149662: gap of unknown length
* 149663 157532: contig of 7870 bp in length
* 157533 157632: gap of unknown length
* 157633 168762: contig of 11130 bp in length
* 168763 168862: gap of unknown length
* 168863 171961: contig of 3099 bp in length
* 171962 172061: gap of unknown length
* 172062 174294: contig of 2233 bp in length.
FEATURES
source
1. 174294
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-180F6"
ORIGIN
Query Match 12.0% Score 40.4; DB 2; Length 174294;
Best Local Similarity 55.8%; Pred. No. 4.9; 61; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 197 ACAGATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATGTAATGCTTCT 256
DB 117449 AAAAGTAGATGCCCTCACTAGGTTTGAACAGAGTATTTTACATGTAAGATTA 117390
QY 257 GATTAAGCCTGTTCAATAATTCCTTTGCAAGCTCTGACTTACCAAGATGCTGCTAC 316
DB 117389 GGTACCTTTTAAATTCATTTCCCAAGCTTACTTCTTAAAAAGAGTCAGTTT 117350
QY 317 AGACAGAAAGATGGCATT 334


```

repeat_region      14235..14272
                    /rpt_family="(CA)n"
repeat_region      14521..14558
                    /rpt_family="(TTA)n"
repeat_region      14935..14963
                    /rpt_family="AT_rich"
repeat_region      15589..15629
                    /rpt_family="(TTTTC)n"
repeat_region      15631..15765
                    /rpt_family="BI_MM"
repeat_region      15919..16063
                    /rpt_family="BI_MM"
repeat_region      16068..16090
                    /rpt_family="(CAAC)n"
repeat_region      16143..16166
                    /rpt_family="(CAA)n"
repeat_region      16604..18631
                    /rpt_family="(A)n"
repeat_region      18696..18735
                    /rpt_family="(A)n"
repeat_region      19686..19753
                    /rpt_family="CT_rich"
repeat_region      22500..22598
                    /rpt_family="(TTTC)n"
repeat_region      22891..22945
                    /rpt_family="CT_rich"
repeat_region      24470..24595
                    /rpt_family="(RSINE1)"
repeat_region      25058..25105
                    /rpt_family="(TATAA)n"
repeat_region      26176..26465
                    /rpt_family="BA4"
repeat_region      26473..26541
                    /rpt_family="(CA)n"
repeat_region      26917..26939
                    /rpt_family="(CAAA)n"
repeat_region      28766..28877
                    /rpt_family="PBID9"
repeat_region      28878..28909
                    /rpt_family="(CAA)n"
repeat_region      29243..29455
                    /rpt_family="B4"
repeat_region      30293..30336
                    /rpt_family="AT_rich"
repeat_region      30557..31009
                    /rpt_family="LI"
repeat_region      31159..31213
                    /rpt_family="LI"
repeat_region      31219..32472
                    /rpt_family="Lx4"
repeat_region      33069..33221
                    /rpt_family="LI_MM"
repeat_region      33470..33470
                    /rpt_family="(CAA)n"
repeat_region      36123..36255
                    /rpt_family="BI"
repeat_region      37393..37422
                    /rpt_family="AT_rich"
repeat_region      40393..40422
                    /rpt_family="(TA)n"
repeat_region      40692..40712
                    /rpt_family="AT_rich"
repeat_region      41263..41297
                    /rpt_family="AT_rich"
repeat_region      42833..42861
                    /rpt_family="AT_rich"
repeat_region      43372..43402
                    /rpt_family="AT_rich"
repeat_region      44537..44680
                    /rpt_family="BI_MM"
repeat_region      44781..44978
                    complement(44781..44978)

```

```

repeat_region      44993..45026
                    /rpt_family="AT_rich"
repeat_region      45864..45939
                    /rpt_family="CT_rich"
repeat_region      45948..45982
                    /rpt_family="(CA)n"
repeat_region      47163..47222
                    /rpt_family="SS"
repeat_region      47927..48187
                    /rpt_family="Lx8"
repeat_region      48211..48294
                    /rpt_family="Lx8"
repeat_region      48808..48990
                    /rpt_family="B3"
repeat_region      50184..50368
                    complement(50184..50368)
repeat_region      50559..50602
                    /rpt_family="B3"
repeat_region      50559..50602
                    /rpt_family="(TAA)n"

```

Query Match 12.0%; Score 40.2; DB 10; Length 201783;
 Best local Similarity 48.9%; Pred.No.5.4; Indels 0; Gaps 0;
 Matches 108; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```

QY      26 TTTTCTACTACAACTCAATTAAGATGAACAGATGATGGTTAGTCACTGTTTAAAG 85
DB      87178 TTTTATAGCTAAATTTGAAAATTTGCTCATTTCTGTATTTGTTATTAATTAAT 87119
QY      86 AAGATTAATAAGATCACTATCATCTTTGAGCAATTAAGGAGGAGAGATTACCAAC 145
DB      87118 CAGACATGTTTCTTTCAGCTCAGAGGATGAGGTTATTGAGAGGAGAGACCACTG 87059
QY      146 AGTGCTTACAGTGAAGAAACAAGTTAAAGTGAAGTCAAGTCAAGTCAAGTCA 205
DB      87058 AGTGAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 86999
QY      206 ATGCCACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 246
DB      86998 GGGAGAGGAGGAGTGAAGGCGGTGCCACATCACTTCA 86958

```

RESULT 13
 CR318618/c 177655 bp DNA linear HTG 01-MAR-2004
 LOCUS
 DEFINITION
 unordereed pieces.
 CR318618
 VERSION
 CR318618.1 GI:44864762
 KEYWORDS
 HTG; HTGS PHASE1.
 SOURCE
 Danio rerio (zebrafish)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 McIay,K.
 Direct Submision
 Submitted (27-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: ZC3917
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 175828 bases at least Q40
 Consensus quality: 176161 bases at least Q30
 Consensus quality: 176379 bases at least Q20

Insert size: 176755; sum-of-contigs
 Insert size: 188044; 8.4% error; agarose-fp
 Quality coverage: 9.57x in Q20 bases; sum-of-contigs Quality
 coverage: 9.27x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
 consists of 10 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

1 57111: contig of 57111 bp in length
* 57112 57211: gap of 100 bp
* 57212 73074: contig of 15863 bp in length
* 73075 73174: gap of 100 bp
* 73175 80256: contig of 7082 bp in length
* 80257 80356: gap of 100 bp
* 80357 89660: contig of 9304 bp in length
* 89661 89760: gap of 100 bp
* 89761 98899: contig of 9139 bp in length
* 98900 105065: contig of 6066 bp in length
* 105066 105165: gap of 100 bp
* 105166 130251: contig of 25086 bp in length
* 130252 130351: gap of 100 bp
* 130352 166939: contig of 36588 bp in length
* 166940 172134: gap of 100 bp
* 167040 172134: contig of 5095 bp in length
* 172135 172234: gap of 100 bp
* 172235 177655: contig of 5421 bp in length.

```

FEATURES

```

source
  1. 177655
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /db_xref="taxon:7955"
    /clone_id="CH211-3917"
    /clone_lib="CHORI-211"
    1. 57111
      /note="assembly fragment:01750"
      fragment_chain:1
      clone_end:SP6
      vector_side:left"
    57212..73074
      /note="assembly fragment:00526"
      fragment_chain:1"
    73175..80256
      /note="assembly fragment:00319"
      fragment_chain:1"
    80357..89660
      /note="assembly fragment:00418"
      fragment_chain:2"
    89761..98899
      /note="assembly fragment:00153"
      fragment_chain:2"
    99000..105065
      /note="assembly fragment:00044"
      fragment_chain:2"
    105166..130251
      /note="assembly fragment:00847"
      fragment_chain:2"
    130352..166939
      /note="assembly fragment:01200"
      fragment_chain:2"
    167040..172134
      /note="assembly fragment:00235"
      fragment_chain:2"
    172235..177655
      /note="assembly fragment:00087"
      fragment_chain:2"

```

ORIGIN

Query Match 11.8%; Score 39.8; DB 2; Length 177655;

Best Local Similarity 52.8%; Pred. No. 6.9;

Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 ATTATTATTTTCATTGATAGCCGGTCTTTTACTACAACTCAATAGATGAACAGATG 60

Db 116279 ATACATTTTATTCACGACGACTCAATTTTAAATATATATATATATATATGATG 1162220

Qy 61 AATGGGTAGTGCCTGTTTAAAGAGAGTAAATAGATCTATCATCTTTGAGCAA 120

Db 116219 AATTTAAAGTGAATATATATATAGAAATGTCACATTTCTATATTTATATAGTCAA 116160

Qy 121 TAAGGAGGAGGATTCGACAACTGCTGCTTAAAGTGA 163

Db 116159 CATACAGTGTCTCAGCATATACAGACACCCCTTACAAATGTA 116117

RESULT 14

BX537337

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Wp: WORMPEP, Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted. Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiyong Bao and Sean Eddy, submitted), and those beginning 'dir' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmark.shtml CH211-222E23 is from a CHORI-211 BAC library VECTOR: pTRABAC2.1.

FEATURES
source Location/Qualifiers
1. 217375
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-222823"
/clone_1ib="CHORI-211"

ORIGIN

Query Match 11.8; Score 39.8; DB 5; Length 217375;
Best Local Similarity 52.8; Pred. No. 6.8;
Matches 86; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATTATATTTCTTCAATAGCGGTTTCTTACACATCAATATAGTGAACAGATG 60
157782 ATTCATTTTATTCACGACGCTCAATTTTAAATATATATATATGATTTAAATG 157841

DB 61 AATGGTTAGTGACTGTTTATTAAGAAGTAAATTAAGTATCATCATTTGAGCAA 120
157842 AATTTAAAGTGAATATATATAGAAATGTGTCAACATTTCTATATTTATTAAGTCAA 157901

QY 121 TAAGGAGGAGAGATTTCAGCAACAGTGTCTTACAGTGA 163
157902 CATACAGTGTCTCAGATATACAGACACCCCTTCAAAATGTA 157944

RESULT 15
AC068698/c 162605 bp DNA 1linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 14 clone RP11-555P23 map 14, WORKING DRAFT
DEFINITION
ACCESSION AC068698
VERSION AC068698
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 162605)
Birken, B., Linton, L., Nuebaum, C. and Lander, E.
Homo sapiens chromosome 14, clone RP11-555P23
Unpublished
2 (bases 1 to 162605)
Birken, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barra, N., Baetien, V., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Birken, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barra, N., Baetien, V., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Dodge, S., Domingo, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 16, 2000 this sequence version replaced gi:7712243.
ALL repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 555_P_23
Center clone name: 555_P_23
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154906 bases at least Q40
Consensus quality: 159118 bases at least Q30
Consensus quality: 160511 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 161405; sum-of-coverage
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 464: contig of 464 bp in length
* 465 564: gap of 100 bp
* 565 2397: contig of 1833 bp in length
* 2398 2497: gap of 100 bp
* 2498 4504: contig of 2007 bp in length
* 4505 4604: gap of 100 bp
* 4605 10523: contig of 5919 bp in length
* 10524 10623: gap of 100 bp
* 10624 18758: contig of 8135 bp in length
* 18759 18858: gap of 100 bp
* 18859 27777: contig of 8919 bp in length
* 27778 27877: gap of 100 bp
* 27878 40462: contig of 12585 bp in length
* 40463 40562: gap of 100 bp
* 40563 53834: contig of 13772 bp in length
* 53835 53934: gap of 100 bp
* 53935 69669: contig of 15935 bp in length
* 69670 69970: gap of 100 bp
* 69970 87514: contig of 17545 bp in length
* 87515 87615: gap of 100 bp
* 87615 103384: contig of 15770 bp in length
* 103385 103484: gap of 100 bp

This Page Blank (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: November 21, 2005, 07:00:52 ; Search time 818.717 Seconds
(without alignments)
5171.686 Million cell updates/sec

Title: US-10-099-663-1_COPY_1115_1626

Perfect score: 512
Sequence: 1 cagtgatctctcatctcat.....gtctgctacagacagaag 512

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
18:	/cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
19:	/cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
20:	/cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
21:	/cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
22:	/cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
23:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq2:*
25:	/cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
26:	/cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
27:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
28:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	512	100.0	2381	17	US-10-099-663-1	Sequence 1, Appl1
2	326	63.7	336	17	US-10-099-663-2	Sequence 2, Appl1
3	43.6	8.5	413	20	US-10-674-124A-2159	Sequence 2159, Ap
4	42.8	8.4	558	13	US-09-925-065A-572321	Sequence 572321,
5	42.8	8.4	679	13	US-09-925-065A-875384	Sequence 875384,

C	6	42.8	8.4	679	13	US-09-925-065A-875385	Sequence 875385,
C	7	42.8	8.4	679	13	US-09-925-065A-875386	Sequence 875386,
C	8	42.4	8.3	5908	15	US-10-239-676-94	Sequence 94, Appl1
C	9	42.4	8.3	5908	16	US-10-240-453-106	Sequence 106, Appl
C	10	42.4	8.3	5908	18	US-10-221-613-176	Sequence 176, App
C	11	42.2	8.2	1238	13	US-09-925-065A-710652	Sequence 710652,
C	12	41.8	8.2	10048	18	US-10-221-613-211	Sequence 211, App
C	13	40.8	8.0	546	13	US-09-925-065A-121976	Sequence 121976,
C	14	40.6	7.9	7040	16	US-10-172-086-13	Sequence 13, Appl
C	15	40.6	7.9	7040	19	US-10-221-714A-161	Sequence 161, Appl
C	16	40.6	7.9	7040	20	US-10-311-507-47	Sequence 47, Appl
C	17	40.6	7.9	7040	21	US-10-480-846-13	Sequence 173, Appl
C	18	40.6	7.9	7040	21	US-10-473-126-173	Sequence 173, App
C	19	40.6	7.9	7040	21	US-10-473-126-319	Sequence 319, App
C	20	40.6	7.9	14095	16	US-10-311-455-450	Sequence 450, App
C	21	40.4	7.9	470	13	US-09-925-065A-178215	Sequence 178215,
C	22	40.4	7.9	470	13	US-09-925-065A-178216	Sequence 178216,
C	23	40.4	7.9	1816	13	US-09-925-065A-45365	Sequence 45365, A
C	24	40.2	7.9	6065	18	US-10-221-613-221	Sequence 221, App
C	25	39.4	7.7	5265	16	US-10-240-453-283	Sequence 293, App
C	26	39.2	7.7	9117	16	US-10-311-455-1042	Sequence 1042, App
C	27	39	7.6	2933	26	US-11-097-143-25165	Sequence 25165, A
C	28	39	7.6	5311	16	US-10-311-455-992	Sequence 992, App
C	29	39	7.6	6202	15	US-10-175-523-189	Sequence 189, App
C	30	39	7.6	6202	26	US-11-099-266-189	Sequence 189, App
C	31	39	7.6	6202	19	US-10-336-603A-35	Sequence 35, Appl
C	32	39	7.6	22052	26	US-11-097-143-28054	Sequence 28054, A
C	33	39	7.6	26097	26	US-11-097-143-2785	Sequence 2785, Ap
C	34	39	7.6	36194	26	US-11-097-143-28114	Sequence 28114, A
C	35	39	7.6	59667	26	US-11-097-143-20479	Sequence 20479, A
C	36	38.8	7.6	1024	21	US-10-425-115-97707	Sequence 97707, A
C	37	38.4	7.5	32392	22	US-10-706-633-27	Sequence 27, Appl
C	38	38.4	7.5	50000	22	US-10-706-635-23	Sequence 23, Appl
C	39	38.2	7.5	6072	16	US-10-311-455-4	Sequence 4, Appl
C	40	38.2	7.5	56737	9	US-09-782-378A-17	Sequence 17, Appl
C	41	38.2	7.5	684973	9	US-09-263-959-1	Sequence 1, Appl
C	42	38	7.4	6802	16	US-10-311-455-1204	Sequence 1204, Ap
C	43	38	7.4	640881	9	US-09-790-988-1	Sequence 1, Appl
C	44	37.8	7.4	3322	26	US-11-097-143-13510	Sequence 13510, A
C	45	37.8	7.4	17527	16	US-10-311-455-1406	Sequence 1406, Ap

ALIGNMENTS

RESULT 1
US-10-099-663-1
; Sequence 1, Application US/10099663
; Publication No. US20030177516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian Gut-Specific Promoters
; FILE REFERENCE: A181
; CURRENT APPLICATION NUMBER: US/10/099,663
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2381
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(1626)
; NAME/KEY: exon
; LOCATION: (1627)..(1693)
; NAME/KEY: Intron
; LOCATION: (1694)..(2322)
; NAME/KEY: exon
; LOCATION: (2333)..(2381)
US-10-099-663-1
Query Match 100.0%; Score 512; DB 17; Length 2381;

Best Local Similarity 100.0%; Pred. No. 8,1e-124;
Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CAGTGATCTTCATCTCATGCTTATTATGTGGAGTAGAATGATCTGCACCAATTAGA 60
DB 1115 CAGTGATCTTCATCTCATGCTTATTATGTGGAGTAGAATGATCTGCACCAATTAGA 1174
OY 61 ATGGACAAGAGAGATTTGTGTTTATCTGTTGGTAAATACGTTTCTCCAGTTGAT 120
DB 1175 ATGGACAAGAGAGATTTGTGTTTATCTGTTGGTAAATACGTTTCTCCAGTTGAT 1234
OY 121 AAAGACCTCCACACAGATTAAGTCTTATGCAACAAGAAATGCAATGATCTCTT 180
DB 1235 AAAGACCTCCACACAGATTAAGTCTTATGCAACAAGAAATGCAATGATCTCTT 1294
OY 181 AGTCTATTATTATTTTCAATGATAGCCGGTTTCTTACCACTCAAAATGAAGTAAC 240
DB 1295 AGTCTATTATTATTTTCAATGATAGCCGGTTTCTTACCACTCAAAATGAAGTAAC 1354
OY 241 AGAATGAATGGGTAGTGACTGTTTATAAGAAAGATTAATTAAGATCATCATTTG 300
DB 1355 AGAATGAATGGGTAGTGACTGTTTATAAGAAAGATTAATTAAGATCATCATTTG 1414
OY 301 AGGCAATTAAGGAGGAGAGATTCAGCAAAAGTGTCTTACAGTGAAGAAACAAGTTAA 360
DB 1415 AGGCAATTAAGGAGGAGAGATTCAGCAAAAGTGTCTTACAGTGAAGAAACAAGTTAA 1474
OY 361 ACTAAAGTACCCCTCTCTTGCAAGATCATGCCAGTGGAGTTTACGCCAGCACA 420
DB 1475 ACTAAAGTACCCCTCTCTTGCAAGATCATGCCAGTGGAGTTTACGCCAGCACA 1534
OY 421 TCATCATGTAATTTGCTTCTTGATAGGCTGTTCAATTAATCTCTTGCAAAAGCTCTGC 480
DB 1535 TCATCATGTAATTTGCTTCTTGATAGGCTGTTCAATTAATCTCTTGCAAAAGCTCTGC 1594
OY 481 TACTTACAGAAAGTCTGCCTACAGACAGAAAG 512
DB 1595 TACTTACAGAAAGTCTGCCTACAGACAGAAAG 1626

RESULT 2
US-10-099-663-2
; Sequence 2, Application US/10099663
; Publication No. US20030177516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian Gut-Specific Promoters
; FILE REFERENCE: A181
; CURRENT APPLICATION NUMBER: US/10/099,663
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(336)
; OTHER INFORMATION: chick intestinal fatty acid binding protein promoter region
US-10-099-663-2
```

Query Match 63.7%; Score 326; DB 17; Length 336;
Best Local Similarity 100.0%; Pred. No. 2.6e-75;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 187 ATTATATTCTTATAGATAGCCGTTTCTTACCACTCAATTAAGATGAAGAGATG 246
DB 1 ATTATATTCTTATAGATAGCCGTTTCTTACCACTCAATTAAGATGAAGAGATG 60
OY 247 AATGGTTAGTGACTGTTTATAAGAAAGATTAATTAAGATCATCATCTTTGAGGCA 306
DB 61 AATGGTTAGTGACTGTTTATAAGAAAGATTAATTAAGATCATCATCTTTGAGGCA 120
```

```
OY 307 TAAGGAGGAGAGATTCAGCAAAAGTGTCTTCAAGTGGAAAACAAGTTAAACTAA 366
DB 121 TAAGGAGGAGAGATTCAGCAAAAGTGTCTTCAAGTGGAAAACAAGTTAAACTAA 180
OY 367 GTACCCCTCTCTTGAAGAATCAATCCACAGTTGAGCTTTAGCCACCATCATCA 426
DB 181 GTACCCCTCTCTTGAAGAATCAATCCACAGTTGAGCTTTAGCCACCATCATCA 240
OY 427 TGTAAATGCTTCTGATTAAGCTGTTCATTAATTTCTTTGCAAAAGCTGCTACTTA 486
DB 241 TGTAAATGCTTCTGATTAAGCTGTTCATTAATTTCTTTGCAAAAGCTGCTACTTA 300
OY 487 CCAGAAAGTGTGCTTACAGACAGAAAG 512
DB 301 CCAGAAAGTGTGCTTACAGACAGAAAG 326
```

```
RESULT 3
US-10-674-124A-2159/c
; Sequence 2159, Application US/10674124A
; Publication No. US2004019797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hideoshi
; APPLICANT: TAMURA, Gen
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; FILE REFERENCE: ORIN-003C1P
; CURRENT APPLICATION NUMBER: US/10/674,124A
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 2159
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: AC012079.4_71311
; FEATURE:
; OTHER INFORMATION: Located on chromosome 1
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; FEATURE:
; OTHER INFORMATION: sequence : 250648904
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chromosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 205277
US-10-674-124A-2159
```

Query Match 8.5%; Score 43.6; DB 20; Length 413;
Best Local Similarity 55.2%; Pred. No. 0.47;
Matches 85; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
OY 138 TATTAAGTCTTATGCAACAAGAAATGCAATTAATCTCTTACTGATTAATTTT 197
DB 256 TATTAAGTCTTATGCAACAAGAAATGCAATTAATCTCTTACTGATTAATTTT 197
OY 198 CATTAAGTACCGGTTTCTTACCACTCAATTAAGATGAAGAGATGAGTAACT 257
DB 196 TATTAATTAATTAATTAATTTGCTCTCACTCAAAAGGCTACAGAAAGCAAGTACTAT 137
OY 258 GACTGTTTATAAGAAAGTAAATTAAGATTAAT 291
```

```

Db      136  GACCTTATAGTAGCAATATAGTGAATCTTAATTT 103

RESULT 4
US-09-925-065A-572321/c
; Sequence 572321, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 572321
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-572321

Query Match      8.4%; Score 42.8; DB 13; Length 558;
Best Local Similarity 53.6%; Pred. No. 0.86;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0

OY      187  ATTATATTATTCATTAGTAGTAGGCGGGTTTTCCTACTCAACTCAATTAAGTAGCAAGATG 246
DB      346  ATTAACATTAACTAAATTTAGTACCCCTTTATATCTCTTTAACACGACACATTTAGGAAC 287
OY      247  AATGGGTAGTAGCAGCTTTATATTAAGAGGTAATAGATCTATCATCATTTGAGGCA 306
DB      286  TAAGGTATAGTCATGTATATTAAGAAATAGATAGAACATTAATTAAGATAGCA 227
OY      307  TAAGGAGAGGAGAGATTTCAGCAAAACAGTGTGCTTCAAGTGGA 352
DB      226  TAAAGAACCTAAATCTTAGAAATTAATTAATAAAATGTCAAA 181

RESULT 5
US-09-925-065A-875384/c
; Sequence 875384, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875384

```

```

; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875384

Query Match
Best Local Similarity 8.4%; Score 42.8; DB 13; Length 679;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

187 ATTATTTATTTTCATTAGATAGCCGGTTTTTACTACAACTCAATAAGATGAACAGATG 246
    |||||
515 ATTAACTTAACAAATTAAGTAGCCTTTATATATCTTTAAACAAGAACATTTAGAGACT 456
OY 247 AATGGGTTAGTGACTGTTTATTAAGAAGAGTAAATGAATATCTATCATTTGAGCAA 306
    455 TAAAGAGTAAGCCATGTATATATAAGAAATTAAGAAACATTAATATAAGAAATTAACAA 396
OY 307 TAAAGGAGGGAGGAGATTACAGAAACAATGTGCTTACAAAGTGAGAAA 352
    395 TAAAGACACTTAATAATCTTAGAAAATAAATTTAATATAAAATGTACAAA 350
DB

RESULT 6
US-09-925-065A-875385/c
; Sequence 875385, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875385
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875385

Query Match
Best Local Similarity 8.4%; Score 42.8; DB 13; Length 679;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

187 ATTATTTATTTTCATTAGATAGCCGGTTTTTACTACAACTCAATAAGATGAACAGATG 246
    |||||
515 ATTAACTTAACAAATTAAGTAGCCTTTATATATCTTTAAACAAGAACATTTAGAGACT 456
OY 247 AATGGGTTAGTGACTGTTTATTAAGAAGAGTAAATGAATATCTATCATTTGAGCAA 306
    455 TAAAGAGTAAGCCATGTATATATAAGAAATTAAGAAACATTAATATAAGAAATTAACAA 396
OY 307 TAAAGGAGGGAGGAGATTACAGAAACAATGTGCTTACAAAGTGAGAAA 352
    395 TAAAGACACTTAATAATCTTAGAAAATAAATTTAATATAAAATGTACAAA 350
DB

RESULT 7
US-09-925-065A-875386/c
; Sequence 875386, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:

```

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 875386
LENGTH: 679
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-875386
```

```
Query Match      8.4%; Score 42.8; DB 13; Length 679;
Best Local Similarity 53.6%; Pred. No. 0.94;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```
QY 187 ATTATATTTTCTAGATAGCGGTTTCTACACTCAATTAAGTGAACGATG 246
    |||||
DB 515 ATTACATTAAACAATTAGTAGCCTTATATCTTTAAACAAGCACTTTAGAGAACT 456
QY 247 AATGGTGTGACCTGTTTAAAGAGACTAATTAAGTATCATTCATTGGAGCA 306
    |||||
DB 455 TAAGAGTAAGTCCATGATTAATAGATTAAGTAAGAACACTATATATAGATAGCAA 396
QY 307 TAAGGAGGAGAGATTCAACAACGCTGCTTACAGTGGAAAA 352
    |||||
DB 395 TAAAGACACTAATACTTGAATAATTAATTAATAAATGTACAA 350
```

```
RESULT 8
US-10-239-676-94/c
; Sequence 94, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10033529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 94
; LENGTH: 5908
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-94
```

```
Query Match      8.3%; Score 42.4; DB 15; Length 5908;
```

```
Best Local Similarity 54.5%; Pred. No. 3.1;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 138 TATTAAGTCCTATGCAACAAGAAATGTCATATCTCTTAGCTCATTTATTTT 197
    |||||
DB 4703 TATACTCTTTTATATTAACAATATATCTATATCTATAAACCCTACTTTTAA 4644
QY 198 CATTAGATAGCGGTTTCTTACTACACTCAATTAAGTGAACAGATGAATGGTTAGT 257
    |||||
DB 4643 TATTATATTTTATTTCTACTACTCTTCTCTAAATTTAAAAAATTAATCTATACCTTAAT 4584
QY 258 GACTGTTTATAAGAGATTAATTAAGATCTATCA 293
    |||||
DB 4583 TCCCTTTATCCAAATTAATTAATCTATATAGCAACA 4548
```

```
RESULT 9
US-10-240-453-106/c
; Sequence 106, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 106
; LENGTH: 5908
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-106
```

```
Query Match      8.3%; Score 42.4; DB 16; Length 5908;
Best Local Similarity 54.5%; Pred. No. 3.1;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

```
QY 138 TATTAAGTCCTATGCAACAAGAAATGTCATATCTCTTAGCTCATTTATTTT 197
    |||||
DB 4703 TATACTCTTTTATATTAACAATATATCTATATCTATAAACCCTACTTTTAA 4644
QY 198 CATTAGATAGCGGTTTCTTACTACACTCAATTAAGTGAACAGATGAATGGTTAGT 257
    |||||
DB 4643 TATTATATTTTATTTCTACTACTCTTCTCTAAATTTAAAAAATTAATCTATACCTTAAT 4584
QY 258 GACTGTTTATAAGAGATTAATTAAGATCTATCA 293
    |||||
DB 4583 TCCCTTTATCCAAATTAATTAATCTATATAGCAACA 4548
```

```
RESULT 10
US-10-221-613-176/c
; Sequence 176, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
```



```

; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10035229.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 176
; LENGTH: 5908
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-176

Query Match
Best Local Similarity 8.3%; Score 42.4; DB 18; Length 5908;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 138 TATAAGTCTCTATGCAACAAGAAATGTCATACCTCTCTAGTCTATTAATTTT 197
DB 4703 TATAACTCTTTTATTAATAAACAATATACATATATCTCTATTAACCACTCTTTTAA 4644
QY 198 CATTAGATAGCCGCTTTTACTACAACTCAATTAAGTGAACAGATGCGTAGT 257
DB 4643 TATTATATTTTATTTTACACACCTTTCTCTAAATTTAAATTAATTAATCCTTAAT 4584
QY 258 GACTGTTTATTAAGAAGAGTAAATTAAGATACATCA 233
DB 4583 TCCCTTTATCCAAATTAATTAATCTATATACGACA 4548

RESULT 11
US-09-925-065A-710652/c
; Sequence 710652, Application US/03925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 710652
; LENGTH: 1238
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-710652
```

```

Query Match
Best Local Similarity 8.2%; Score 42.2; DB 13; Length 1238;
Matches 92; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 74 AGATTGCTTTTATCTGTTGGTAAATAGTTTCTCCAGTTGTATAAGACCTCCCA 133
DB 986 ATATATGAGGTTCTTGGCTCAGCATATCAATTTATGCAATGATGAACACATACAA 927
QY 134 CCAATTAAGTCCATATGCAACAAGAAATGTCATATACATTCCTTATCTCATTTA 193
DB 926 ATCTCATGAATGAATGATGATACATATAGTGCCTCAATGACATGTTTCTATGATTT 867
QY 194 TTTTATGATAGCCGCTTTTCTAATCTCAATTAAGTGAACAGATGAA 248
DB 866 CTATATATTAATAAGATATTTTGTGTAATTCAGAAATCTAATAACATTTGAA 812

RESULT 12
US-10-221-613-211/c
; Sequence 211, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10035229.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 211
; LENGTH: 10048
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3184, 3255, 3396, 6016, 8568)
US-10-221-613-211

Query Match
Best Local Similarity 8.2%; Score 41.8; DB 18; Length 10048;
Matches 136; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 100 ATACGTTTCTCCAGTTGTATAAGACCTCCCAAGATTAAGCTATCAACAAG 159
DB 5471 ATATTTATTTATCTTAAACCTAAATTCCTCTCTACAAAAAATAATTAATCAATAC 5412
QY 160 AAATGTCATATCATTCCTTATGCTCATATTAATTTTATTAAGATGACGCTTTTAC 219
DB 5411 TCCAAATCCAAAAATCTATTTTAAACCTATTAATAATCCCTCAATAATCACTTCTAC 5352
QY 220 TACAATCAATAAATGAAGATGAATGATGATGATGATGATGATGATGATGATGATG 279
DB 5351 AATATCAATAAATAAATAAACAACATTAATTCCTTAATCTAATAATCAATAAATAA 5292
QY 280 TAAAGATCAATCATTTGAGCAATTAAGAGAGAGAGAGATTCAGCAACAGTGTGCT 339
DB 5291 TATTAATTTACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTCA 5232
```

QY 340 TACAGTGGAAAAAGATTAAAGTACCCCTCTTACAGATCA 392
DB 5231 AACAACAAACAAACAAACAAACAAACCAATTTCAACCTATCTA 5179

RESULT 13

US-09-925-065A-121976
Sequence 121976, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: PASTESEQ for Windows Version 4.0
SEQ ID NO 121976
LENGTH: 546
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-121976

Query Match
Best Local Similarity 8.0%; Score 40.8; DB 13; Length 546;
Matches 93; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

QY 106 TTCTCCAGTTGTATTAAGACCTCCACAGATTAAGTCCATGCCAACAAGAAAG 165
DB 104 TATGTAAAGTGTATGATGCTGAGGTAGACATGAGTAAGAAAAATATTAAT 163
QY 166 TCAATACATCTCTCTAGTCTCATTTATTTTCATTAGATAGCCGTTTTTACTACAAC 225
DB 164 TTAATTAATTTCTGCTGGAACATGGTTATTTATATGAAAATATAGTAACATTTCTTA 223
QY 226 TCAATTAAGATGAACAGATGAATGGCTTAGTGCATGTTTATTAAGAGATTAATAAGA 285
DB 224 CATCATAGACAGTAATGAATGTACGATGTATGATATTAACAATTAGAAAATAAGA 283
QY 286 TA 287
DB 284 GA 285

RESULT 14

US-10-172-086-13/C
Sequence 13, Application US/10172086
Publication No. US20030113750A1
GENERAL INFORMATION:

APPLICANT: Epigenomics AG
TITLE OF INVENTION: Method and nucleic acids for the differentiation
of prostate tumors
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/172,086
PRIOR FILING DATE: 2002-06-13
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 13
LENGTH: 7040
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-172-086-13

Query Match
Best Local Similarity 7.9%; Score 40.6; DB 16; Length 7040;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 101 TACGTTTTCTCAGTGTATTAAGACCTCCACAGATTAAGTCTATGCAACAAAGA 160
DB 3586 TATATTTTAAAAAATCAATTAATTAATTCCTTCTTCAAAACCTTATATTTAAAT 3527
QY 161 AATGTCATATCAATCTCTTACTCATTTATTTTCATTAGATAGCCGTTTTTACT 220
DB 3526 ATATATTAACCATTTTATATATTAACAATTCATTTTAAAAATATTAACCTTAAAAA 3467
QY 221 ACACTCAATTAAGATGAACAGATGAATGGTTAGTACTGTTTATTAAGAAGTAT 280
DB 3466 ATAAATTAATAACATTAACCAATTTACAAAAACAATTTAAATCTTAACCCCTAAAT 3407
QY 281 AAAGATCTAT 291
DB 3406 AAAATTTCTAT 3396

RESULT 15

US-10-221-714A-161/C
Sequence 161, Application US/10221714A
Publication No. US20040048254A1
GENERAL INFORMATION:

APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
tumor suppressor genes and oncogenes
FILE REFERENCE: 5013.1005
CURRENT APPLICATION NUMBER: US/10/221,714A
PRIOR FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: PCT/EP01/02955
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: DE 10013847.0
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 161
LENGTH: 7040
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-161

Query Match
Best Local Similarity 7.9%; Score 40.6; DB 19; Length 7040;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 101 TACGTTTTCTCAGTGTATTAAGACCTCCACAGATTAAGTCTATGCAACAAAGA 160
DB 3586 TATATTTTAAAAAATCAATTAATTAATTCCTTCTTCAAAACCTTATATTTAAAT 3527
QY 161 AATGTCATATCAATCTCTTACTCATTTATTTTCATTAGATAGCCGTTTTTACT 220
DB 3526 ATATATTAACCATTTTATATATTAACAATTCATTTTAAAAATATTAACCTTAAAAA 3467
QY 221 ACACTCAATTAAGATGAACAGATGAATGGTTAGTACTGTTTATTAAGAAGTAT 280
DB 3466 ATAAATTAATAACATTAACCAATTTACAAAAACAATTTAAATCTTAACCCCTAAAT 3407

Qy	281	AAAGATACCTAT	291
Db	3406	AAATTCTAT	3396

Search completed: November 21, 2005, 11:54:58
Job time : 822:717 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2005, 05:12:21 ; Search time 126.189 Seconds
(without alignments)
6639.056 Million cell updates/sec

Title: US-10-099-663-1_COPY_1115_1626

Perfect score: 512
Sequence: 1 cagtgatgctctcatctcatc.....gtctgctacagacagaag 512

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCUS COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.4	8.9	1141	US-09-806-708B-22	Sequence 22, Appl
2	42.2	8.2	1257	US-09-270-767-15058	Sequence 15058, A
3	40	7.8	733	US-09-270-767-10136	Sequence 10136, A
4	39.2	7.7	4860	US-09-270-767-10504	Sequence 10504, A
5	38	7.4	640681	US-09-790-988-1	Sequence 1, Appl1
6	37.8	7.4	7218	US-08-232-463-14	Sequence 14, Appl1
7	37.6	7.3	601	US-09-949-016-125489	Sequence 125489, A
8	37.6	7.3	44353	US-09-949-016-15302	Sequence 15302, A
9	37.4	7.3	663	US-08-998-416-187	Sequence 187, App
10	37.4	7.3	696	US-08-998-416-779	Sequence 779, App
11	37.4	7.3	719	US-08-998-416-1138	Sequence 1138, App
12	37.4	7.3	856	US-08-998-416-289	Sequence 289, App
13	37	7.2	832	US-09-621-976-2813	Sequence 2813, App
14	36.8	7.2	43117	US-09-949-016-17589	Sequence 17589, A
15	36.6	7.1	601	US-09-949-016-165407	Sequence 165407, A
16	36.6	7.1	601	US-09-949-016-165464	Sequence 165464, A
17	36.6	7.1	27543	US-09-949-016-17455	Sequence 17455, A
18	36.6	7.1	36643	US-09-949-016-11860	Sequence 11860, A
19	36.6	7.1	36821	US-09-949-016-16403	Sequence 16403, A
20	36.6	7.1	36821	US-09-949-016-16404	Sequence 16404, A
21	36.2	7.1	601	US-09-949-016-173718	Sequence 173718, A
22	36.2	7.1	601	US-09-949-016-173765	Sequence 173765, A
23	36.2	7.1	832	US-09-621-976-2813	Sequence 2813, App
24	36.2	7.1	1587	US-09-710-279-1929	Sequence 1929, App
25	36.2	7.1	1878	US-09-134-001C-1667	Sequence 1667, App
26	36.2	7.1	3000	US-09-710-279-3907	Sequence 3907, App
27	36.2	7.1	3368	US-09-710-279-4169	Sequence 4169, App

28	36.2	7.1	81819	US-09-949-016-16661	Sequence 16661, A
29	36.2	7.1	81819	US-09-949-016-16662	Sequence 16662, A
30	36	7.0	1055	US-09-806-708B-23	Sequence 23, Appl
31	36	7.0	168394	US-09-949-016-13002	Sequence 13002, A
32	36	7.0	786431	US-09-751-389-3	Sequence 3, Appl1
33	35.8	7.0	1557	US-09-270-767-28519	Sequence 28519, A
34	35.8	7.0	3128	US-09-270-767-12706	Sequence 12706, A
35	35.8	7.0	192956	US-09-949-016-14382	Sequence 14382, A
36	35.8	7.0	1230025	US-09-438-452A-1	Sequence 1, Appl1
37	35.8	7.0	1230230	US-09-438-185A-1	Sequence 1, Appl1
38	35.6	7.0	601	US-09-949-016-44395	Sequence 44395, A
39	35.6	7.0	601	US-09-949-016-58898	Sequence 58898, A
40	35.6	7.0	1137	US-09-107-532A-2846	Sequence 2846, App
41	35.6	7.0	139562	US-09-949-016-13451	Sequence 13451, A
42	35.6	7.0	256171	US-09-949-016-12822	Sequence 12822, A
43	35.6	7.0	256176	US-09-949-016-15524	Sequence 15524, A
44	35.4	6.9	601	US-09-949-016-121946	Sequence 121946, A
45	35.4	6.9	601	US-09-949-016-202243	Sequence 202243, A

ALIGNMENTS

```
RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806, 708B
; PRIOR APPLICATION NUMBER: 2001-04-03
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.c., L.a., and B.n. PAEI promoters
US-09-806-708B-22

Query Match      8.9%; Score 45.4; DB 4; Length 1141;
Best Local Similarity 9.9%; Pred.No. 0.003;
Matches 46; Conservative 197; Mismatches 221; Indels 0; Gaps 0;

QY 12 TCATCTCATGCTTATATGAGAGTATGATTCATCCAAATAGATGAGCAAGC 71
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 110 YBMTWTKKMGKQWRRYRMRABDVTVDHYTANNNATWTCMDXDKRTRWKK 169
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 72 AGAGATTTGTTTATCTGTTGGTAAATACGTTTCCAGTTTATTAAGACCTCC 131
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 170 NNATGDDDTKXMMNNNCCBTVTWVVRXYKTRDWBKXMMNGMBWKKWSYDVTY 229
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 132 CACCGATATAAGTCTATGCAACAAGAAATGCAATACATTCCTAGTCTCTAT 191
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 230 WDMCRKRVRRVRRGRMRNTVAVBTYHRRRRYNGWBTAAVYRWYNNNNNNKANC 289
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 192 TATTTTCATTAGTAGCCGTTTTCATACACCTCAATTAAGATGAGCAATGATG 251
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 290 KRAKVGWRABVNSCTTWKSKTKTYRSCVANNRABDAKDKHKKWKSAAAGYWN 349
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 252 GTTAGTACTGTTTATTAAGAGAGTAAATGAATACATTCATTCATTGAGCAAT 311
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 350 NNNNNNTYKKBABRMDVWMSAWKMHANAAHRSKKQVBYKRTVNNNNNGTMMKR 409
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 312 GAGGAGAGATTCAGCAAAAGTGTGCTTCAAGTGAAGAAACAGTAAACTGAATG 371
    :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Db 410 MMAMYKMDMDHBTGTYNNNNNGRTYYGWTIKKKMMWYTKKXANNCKRAMPDHKTCTHNN 469
Qy 372 CCCCCCTCTGACAAATCAATGCCAGTTGAGCTTTAGCCAGCCATCATCATGTA 431
Db 470 TTMWKKTKTYNNNNYKSMNKGSHRBAALVYTWMMWRYYAANNNDYWKACTWYK 529
Qy 432 ATTCCTTCCTGTAAGCCTGTGTCATTAATCTCTTTGCAAGC 475
Db 530 YBVCSTKMMNNYAAWYTKSSMNTSRIRYKTKNNSMWRSDTSM 573

RESULT 2
US-09-270-767-15058/c
; Sequence 15058, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 15058
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-15058

Query Match 8.2%; Score 42.2; DB 4; Length 1257;
Best Local Similarity 53.3%; Pred. No. 0.026;
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 73 GAAATTTGCTTTATCTGTGGTAAATACGTTTCTCCAGTTGATATAAGCCTCCC 132
Db 364 GAATTTATGATTAATTAATTTAGTTATTTCACTCAACACGACGACGACTACTTC 305
Qy 133 ACCAGTATAAGTCTATGCAACAAAGAAATGTCATATCATCTCTTAGTCTATT 192
Db 304 ACAGCAAAACGTCACAGAGGAAAGAAAGAAATTAAGAGTGATTTCTCTCACAATA 245
Qy 193 ATTTGATAGATAGCGGTTTCTTACTCACTCAATAAGATGA 239
Db 244 TGTTTATGGCATAAAGGTGTGGCATTCATCAATAATAAGTA 198

RESULT 3
US-09-270-767-10136/c
; Sequence 10136, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10136
; LENGTH: 733
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-10136

Query Match 7.8%; Score 40; DB 4; Length 733;
Best Local Similarity 50.5%; Pred. No. 0.092;
Matches 97; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 48 TCACCAATTAAGATGACAAAGCAGAGATTGTGTTTATCTGTGGGTAATAGCTTT 107
Db 306 TAAATTAAGTCAAGTTTATTTATTTAGTTTATGATTTTATTTTAAAGTATTTCACCT 247
Qy 108 TCTCAGTTGTATAAGACCTCCACAGTATTAAGTCTATGCAACAAAGAAATGTC 167

Db 246 GCAACACGACACCGACGACCTACTCAACAGCAAAAAAGTACAGAGAAAGAAAGATTA 187
Qy 168 AATACATTTCTTACTGTCATTAATTTATTTTCAATGATAGCCGTTTTTTACTACACTG 227
Db 186 AAAAAAGGTATCTCTTCAATATGTTTATGCAATAAAGGTGGCCATTGATATC 127
Qy 228 AATATAGATGA 239
Db 126 AATATTAAGTA 115

RESULT 4
US-09-270-767-10504/c
; Sequence 10504, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10504
; LENGTH: 4860
; TYPE: DNA
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-10504

Query Match 7.7%; Score 39.2; DB 4; Length 4860;
Best Local Similarity 46.6%; Pred. No. 0.33;
Matches 125; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 97 TAAATACGTTTCTCAGTTGATTAAGACCTCCACGATTAAGTCTATGCAACA 156
Db 1390 TAAATCAGATTAATTTGTATTAATATCACACGATTAATTAATTTGAAGCATTAATT 1331
Qy 157 AAGAAATGTCATATCTCTTAGTCTCATTAATTTTTCATTAATGATAGCCGTTTTT 216
Db 1330 AATTAATTAATTAATTAATTAATGAGCCATATGCTTTAAATTTTTCGCTTCACA 1271
Qy 217 TACTCACTCAATTAAGATGAACAAATGAATGAGTATGATGCTTTTAAAGAG 276
Db 1270 TTTTGACGAATTAATTTGTTAAATTTTACATTTGATGATCTTTTAAAGAGC 1211
Qy 277 TAATTAAGTACTATCATTAATTTGAGCAATTAAGGAGGAGAGATTCAGCAACAGTGT 336
Db 1210 CAAATTAAGTTTATCTTCAATATGAGTAAGAAAGATTAATTAATTAACCAATTAGTTC 1151
Qy 337 GCTTACAGTGAAGAAACAAATTAACCTA 364
Db 1150 ATTTTATGTTAAATTAAGTCAATGTCGA 1123

RESULT 5
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: MATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-1071160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1

LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match
Best Local Similarity 46.6%; Pred. No. 4.9;
Matches 122; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

7.4%; Score 38; DB 4; Length 640681;
Query 24 TATTATGTGAGTAGAATGATGCTCAACAAATTAGAAGCAAGAGATTGTGT 83
Db 189923 TATTATGCTGAGAGAAATGATGATGATTTCTCAAAATTTGAATCGAGATTAATTT 189982
QY 84 TTTATGCTGGGTAATAGTTTCTCCAGTGTATTAAGACCCCTCCACGATATAA 143
Db 189983 TATAGCAAAATTTGAAAAATCTCTCCGATATATTTAAAAATMAAAAAATTTAA 190042
QY 144 GTCCATGACAAAGAAATGTCATACATCTCTAGTCTCATTTATTTTCATTAG 203
Db 190043 TTTTCTTTTAAACAAAAATATTAATAATTTTGGTGTATTTTATCAAAATTA 190102
QY 204 ATAGCCGTTTTTTTACTACAACTCAATTAAGATGACAGATGATGGTTAGTGTG 263
Db 190103 TTCTCCTTTCTAGAAAAACAATATGATTTTATTAAGTTGATGTTTAAAAAAT 190162
QY 264 TTTTAAAGAGAGTAAATGA 285
Db 190163 AAAAAATGAAACAAACAAATA 190184

RESULT 6

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232.463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)863-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpdt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 13.5%; Pred. No. 0.97;
Matches 30; Conservative 103; Mismatches 90; Indels 0; Gaps 0;

7.4%; Score 37.8; DB 1; Length 7218;
Query 148 TATGCAACAAAGAAATGTCATATCTCTAGTCTCATTTATTTTCATTAGTAG 207
Db 1494 TAGGACTACTGTAATTAATCTATCTGCAAGTAGTTAAGATTAAGAAATTTGGTAC 1435
QY 208 CCGGTTTTTACTACAACTCAATTAAGATGAAGATGAATGGTTAGTGTAT 267
Db 1434 RRR 1375
QY 268 AAGAAGAGATTAAGATATCTATCATTTAGGCAATTAAGGAGGAGATTCAGC 327
Db 1374 RRR 1315
QY 328 AAGAGTGTCTTACAGATGAAGAAACAGTTAACTAAAGTGA 370
Db 1314 RRR 1272

RESULT 7

US-09-949-016-125489/c
Sequence 125489, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CI001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FaersSeq for Windows Version 4.0
SEQ ID NO 125489
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-125489

Query Match
Best Local Similarity 49.0%; Pred. No. 0.42;
Matches 100; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

7.3%; Score 37.6; DB 4; Length 601;
Query 148 TATGCAACAAAGAAATGTCATATCTCTAGTCTCATTTATTTTCATTAGTAG 207
Db 365 TATTAATAATAGATATATATACACAGATATATTAACATGATTAAGTTAACTACATAT 306
QY 208 CCGGTTTTTACTACAACTCAATTAAGATGAAGATGAATGGTTAGTGTAT 267
Db 305 ACATRTATATAGTAGTATATCTAGAAAGCATGCAAGAAATTTCTGCTACTATCTAA 246
QY 268 AAGAAGAGATTAAGATATCTATCATTTAGGCAATTAAGGAGGAGATTCAGC 327
Db 245 TTAATGACATCTTATAGTCTCTCAAAATTTAAGATTAATAATTTCAAGCGTCAGT 186
QY 328 AAGAGTGTCTTACAGATGAAGAA 351
Db 185 CAACATTATGCTTAAAGTAAAAA 162

```

RESULT 8
US-09-949-016-15302/C
Sequence 15302, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO010307
CURRENT FILING DATE: US/09/949, 016
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15302
LENGTH: 44353
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)_(44353)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15302

```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 779:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1501RP
US-08-998-416-779

Query Match
Best Local Similarity 58.6%; Score 37.4; DB 3; Length 696;
Pred. No. 0.51;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 180 TAGTCATTATTATTTCTTATGATGACCGGTTTTTCTACTCACTCAATAGATGAA 239
DB 453 TAGATTATTATTAACTTTATTTGTTAAACATTATTATTATGATCATTAATATATAA 512

QY 240 CAGATGAATGGGTTAGTACTGTTTATTAAGAAGATATAAGTACTA 290
DB 513 GGAATTAACATTATGATATATATAGTTATTATAGAACCAATGAGATACTA 563

RESULT 11
US-08-998-416-1138/c
Sequence 1138, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtel, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1138:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692UP
US-08-998-416-1138

Query Match
Best Local Similarity 58.6%; Score 37.4; DB 3; Length 719;
Pred. No. 0.51;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 180 TAGTCATTATTATTTCTTATGATGACCGGTTTTTCTACTCACTCAATAGATGAA 239
DB 574 TAGATTATTATTAACTTTATTTGTTAAACATTATTATTATGATCATTAATATATAA 515

QY 240 CAGATGAATGGGTTAGTACTGTTTATTAAGAAGATATAAGTACTA 290
DB 514 GGAATTAACATTATGATATATATAGTTATTATAGAACCAATGAGATACTA 464

RESULT 12
US-08-998-416-289/c
Sequence 289, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtel, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 289:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241UP
US-08-998-416-289

```

Query Match	7.3%	Score	37.4	DB	3	Length	856
Best Local Similarity	58.6%	Pred. No.	0.55				
Matches	65	Conservative	0	Mismatches	46	Indels	0
						Gaps	0

Qy	180	TAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Db	574	TAG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

Qy		240 CAGATGAATGGGTAGTACTGTTATTAAAGAAGAGTAATAAAGTACTA	290
Db		514 GGAATACATTATGATATATATAGTATTATTAGAACCAATGAAGTACTA	464

RESULT 13
US-09-621-976-2813

Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.

APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins
FILE REFERENCE: GENSET.054PR2

```

/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm

```

```

; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
;

```

```

; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

```

Query Match	7.2%	Score 37	DB 4	Length 832
Best Local Similarity	15.2%	Pred. No. 0.71		
Matches	45	Conservative	123	Mismatches 123; Indels 1; Gaps 1

```

QY      3 GTGATCCTTCATCTCATGCTTATTATGTGGAGTAGATAGATTCTCAACCAATTAGAAAT 62
      ::: : : ::: : : :| | | | | : : : : : : : : : : : :|
Db     79 KKKKKKGYMMMMYYGWRRSYNAAMTTRIMTGYAYYRSMMYWMWRCKKKKAYRKTTCYSS 138

```

```

Oy      63 GGACAAAGCAGAGATTGTGTTTT-ATCTGTTGGGTAATACTGTTTCTCCAGTTGTATA 121
        :: :: : | :: :: :: | : : | :: :: : | :: :: :
Db     139 KGATWKRWRKKAAWTTWWMKKTYMAATRYMMNMCMTKRMBASWYCWMMGKAIAKMSYTRK 198

```

```

Oy      122  AAGACCTCCCAACAGTATTAAGTCCATATGCACAAAGAAATGTCAATACATCTCTTA 181
        ::  :  :  ||  :  :  ::  :  :  :  :  :  :  :  :  :  :  :  :
Db      199  SRSYASARSACRCCYSGSWGAMSWKYMIMRMFMWRGMAITGAGMCAWRASCMRMRRKYAGSK 256

```

```

Oy      182  GTCCTATTATTTCATTGATAGCCGGITTTTCTACAACCTAAATAGATGAA 241
          ::::  |: :: |::| ::  :  |:: ::::  |::
Db      259  TSYKMMCMCTRSKRCYCTKARMTGYCYRKGMGKRGRTAYSKVYMKRWMCNARY 318

```

```

QY      242 GAATGAATGGGTAGTACTGTTTATAAAGAGAGTATAAAGATCTATCATCATTT 298
        : || : : : : : : : : : : : : : : : : : : : : : : : : :
Db      319 RYSTGTPASMMWRMYTMMKKWKYAAARAAMRWMAAMWMAWRACAAATATATATT 375

```

RESULT 14
US-09-949-016-17589/c

```

;
;
; GENERAL INFORMATION:
;
; APPLICANT: VENTER, J. Craig et al.
;
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949, 016

PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012

```

:
:
: SEQ ID NO 17589
:
: LENGTH: 43117
:
: TYPE: DNA
:

```

US-09-949-016-17589
Query Match

Matches	80	Conservative	0	Mismatches	72	Indels	0	Gaps	0
QY	216	TTACTACAACTCAATAAGATGAACAGAAATGAATGGTTAGTGA	CTGTTTATAAAGAGA	275					

Db 20807 TGAACAATAATATGCTGATGAAATTAATTGGAGGATTAAATAAGTAGATAAATGCA 207
Qy 276 GTATTAAGATCTATCATCTTTGGGCATTAAGGAGGAGAGATTACCAACAAGTG 335

Db 20747 CCTTAAGACAAATAGTAACTAGGAAATAGCCTCTGCGGGAAATCTGAAGACAGGA 20688
Ox 336 TGCCTACAGTGGAAAAACAGTTAACTAAG 367

Db 20687 GTTTTAAAGATGACAACTTAGAATGAAG 20656

US-09-949-016-165407
; Sequence 165407, Application US/09949016
; Patent No. 6812339

1 APPLICANT: VENTER, J. Craig et al.
2
3 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
4
5 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755

PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165407
```

TYPE: DNA
ORGANISM: Human
US-09-949-016-165407

Query Match	7.1%	Score 36.6	DB 4	Length 601
Best Local Similarity	48.3%	Pred No. 0.82		
Matches 131, Conservative	0	Mismatches 139	Indels 1	Gaps 1

QY 134 CCAGTATAAAGTCCTATGCAACAAAGAAATGCAATACATTCCTTGTAGTCATTATTA 193

DB 248 CCAGCCAAACCAATTGACACAAAGAAAGAAATATGTAGCAACTGTATATTATVGGATTT 307

OY 194 TTTTCATTAGATAGCCGGTTTTTCTACTACAACTCAAAATAAGTAGAACGATGAATGGGT 25
| | | | |
Db 308 TTTTAAAAAATTAAAGATTATGTGCATTAATAATGGAGTCTATATGAGAGTGCAA 36

254 TAGTACTGTTTATAAGAAGTATAAAGTACTATCATCATTTGAGGCATAAGGA 311

Db	368		GAGTCTTCATTAAGAAATCATCATATTAATAC-ACCACCATTTAGAAAGAAAGGA	426
Oy	314		GGGAGAGATTGAGCAACAGTGTGCTTACCAAGTGAAAAAAGTTAAACTTAAAGTACCC	373
Db	427		AAATCTAACAGATGACCATCTGGAATAGACGTGAGAAAGCATAGACAAATTCATTA	486
Oy	374		CCCTCCTTGACAAGATCATGCGACAGTTGA	404
Db	487		CTTTTCCTGATTAATTCATCTCAAGATATGA	517

Search completed: November 21, 2005, 08:29:54
Job time : 131.189 secs

Blank (uspto)

may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering_information.htm).

Query Match 9.3%; Score 47.6; DB 9; Length 1206;
Best Local Similarity 45.5%; Pred. No. 0.16;
Matches 170; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

11 TTGATCGAAGCTTATATGAGTAGATATGATCTGACCAATTAAGTAGACAAG 70
DB TTTAACTATATATATTTTATATATATATATATATATATATATATATATG 940
QY 71 CAGAGATTGTTGTTTATCTGTTGAGTAATAGCTTTCCAGTTGTAAGACCTC 130
DB 939 CTAT 880
QY 131 CACACGATATTAAGTCCATATGACAAAGAAAGTCAATATCTCTTACTGATTA 190
DB 879 TAAAT 820
QY 191 TTTATTTTCAATATAGCCGTTTCTTACACACTCAATATAGATGAAGATGA 250
DB 819 ATTAT 760
QY 251 GGTATGACTGTTTATTAAGAGATTAATAGATATATATATATATATATAT 310
DB 759 TATATGATTTTCATATTAACAAATATATATATATATATATATATATAT 700
QY 311 GAGGAGAGATTCAGCAACAGCTGCTTCAAGTGAAGAAACAGTTAACTAA 370
DB 699 CAATTTAT 640
QY 371 CCCCCCTCTTAC 384
DB 639 CACACTTCACTTAC 626

RESULT 2
CNS014PQ 987 bp DNA linear GSS 26-JUL-1999
LOCUS CNS014PQ
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN12P22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL104456
VERSION AL104456.1 GI:5616067
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 987)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Fayan. It has been constructed in the vector
pBelobAC11.

FEATURES
source
1. .987
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN12P22"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"

/note="end : SP6"

Query Match 9.0%; Score 46; DB 9; Length 987;
Best Local Similarity 33.5%; Pred. No. 0.4;
Matches 78; Conservative 51; Mismatches 104; Indels 0; Gaps 0;

QY 136 AGTATTAAGTCTATAGCAAGAAAGATGCAATATCTCTTATAGTCTATATAT 195
DB 723 AAAAAAAAAAAAAAAAAATTTTAAAGAAATTTATTTAAATTTTAAATTTTAA 782
QY 196 TTCTATGATGACCGGTTTTTCTACTCAACTCAATTAAGATGAACGATGCGTTA 255
DB 783 TTAT 842
QY 256 GTGACTGTTTATTAAGAGATTAATATATATATATATATATATATATAT 315
DB 843 TTTTATATATATATTTTAAATATATATATATATATATATATATATATAT 902
QY 316 GAGATTCAGCAACAGCTGCTTACAAAGTGAAGAAACAGTTAACTAACT 368
DB 903 RAATAAT 955

RESULT 3
CNS0080E/c 1101 bp DNA linear GSS 04-JUN-1999
LOCUS CNS0080E/c
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR28A22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL067626
VERSION AL067626.1 GI:4948726
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Oosagawa and
Aaron Mammoxer in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
SmaI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw ap, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28A22"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN
Query Match 8.8%; Score 45.2; DB 9; Length 1101;
Best Local Similarity 41.8%; Pred. No. 0.67;
Matches 79; Conservative 28; Mismatches 82; Indels 0; Gaps 0;

Mon Nov 21 12:27:20 2005

RESULT 4	
AZ541558/c	
LOCUS	AZ541558 910 bp DNA linear GSS 14-NOV-2000
DEFINITION	ENTEB155Tf Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.
ACCESSION	AZ541558
VERSION	AZ541558.1
KEYWORDS	GI:11149198
SOURCE	GSS.
ORGANISM	Entamoeba histolytica
REFERENCE	Entamoeba histolytica
AUTHORS	Eukaryota; Entamoebidae; Entamoeba.
TITLE	1 (bases 1 to 910)
JOURNAL	Loftus, B., Van Aken, S. and Fraser, C.
COMMENT	Determination of clone end sequences from Entamoeba histolytica HM:IMS sheared DNA library Unpublished (2000) Contact: Brendan J Loftus

JOURNAL
Unpublished (2000)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the *Entamoeba histolytica* HMI-1MS8 shared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 26
High quality sequence stop: 840.
location/Qualifiers
1. .910
FEATURES
SOURCE

```

FEATURES
      source                Location/Qualifiers
    1..910
        /organism="Entamoeba histolytica"
        /mol_type="genomic DNA"
        /strain="HMI-1MS8"
        /db_xref="taxon:5759"
        /clone_lib="Entamoeba histolytica Sheared DNA"
        /note="vector: pHDSt; Site_1: Bst I; Constructed at The
              Institute for Genomic Research (TIGR), Rockville, MD.
              Genomic DNA isolated from broth cultures of E. histolytica
              using a method described by Clark and Diamond (Clark,
              C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
              method for isolate identification. Exp. Parasitol.
              77:450.). The DNA was mechanically sheared to give a
              tight size distribution (~2 kb). The v + i method used for
              the library construction is described in detail in Smith,
              H.O. and Venter, J.C. (Making small insert libraries for
              whole genome shotgun sequencing projects. In Genome
              Sequencing: A Practical Approach, eds M. Vaubin and B.
              Barrell, Oxford University Press, 1999)."

```

ORIGIN

Query Match	8.7%	Score 44.4;	DB 8;	Length 910;
Best Local Similarity	46.7%;	Pred. No. 1;		
Matches 141;	Conservative 0;	Mismatches 161;	Indels 0;	Gaps 0;

37 AGATAGATTCTCACCAATTAGATGGACCAAGCAGAGATTGTGTTTATCTGTTGGG 96

Db 310 AGAATTAATAAGAAATAGATTAATTAAGATAAACAATTATTTGTATTTTCATTTGAAT 251

Qy 97 TAAATAGCTTTTCCAGATTGTAATAAGACCCTCCACAGATATAAAGTCTATGCACA 156

Db 250 TAATTCATTATGAATTAATAATGAATATAGATTAATAGAAATTAAGATATAACAATTAAGAAAT 191

Qy 157 AAGAAATAGTCAATCATCTCTTAGTCTCATTTATTTTATTAGATAGCGGTGTTT 216

Db 190 GATAAAATACATCAACATATTTATGTTATTTCTTATTTCTTTATTCATCTTATTTTC 131

Qy 217 TACTTCAACTCAATAATAGATGAACGAATGAATGGGTTAGTACTGTTTATATAAGAAAG 276

Db 130 TATTTAAACAATTAATTAACAACAACATTAATMAAAGATTAATAGATATACACATCATTA 71

Qy 277 TAATTAAGATATCTATCATCTTTGAGGCATAATGAGGAGGAGAGATTCAAGCAACGTCT 336

Db 70 AATAAAGATTAAGATTAATAATATATACAACAACATATATAACAAGAATAATCCAAATACTTT 11

Qy 337 GC 338

Db 10 CC 9

RESULT 5	BH155022	894 bp	DNA	linear	GSS 24-SEP-2001
LOCUS	BH155022				
DEFINITION	BH155022				
	ENT15157F Entamoeba histolytica Sheared				
	genomic, genomic survey sequence.				
ACCESSION	BH155022				
VERSION	BH155022.1				
KEYWORDS	GI:15726739				
SOURCE	GSS.				
ORGANISM	Entamoeba histolytica				
	Entamoeba histolytica				
	Eukaryota; Entamoebidae; Entamoeba.				
REFERENCE	1 (bases 1 to 894)				
AUTHORS	Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.				
TITLE	Determination of clone end sequences from Entamoeba histolytica				
	HMI:IMSS sheared DNA library (2001)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Brendan J Loftus				

JOURNAL
 COMMENT
 Unpublished (2001)
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel.: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@icigr.org
 Clones are derived from the *Entamoeba histolytica* Hm1.IMSS shared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 52
 High quality sequence stop: 822.
 Location/Qualifiers
 1..894
 source

```

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHS01, Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Sarell, Oxford University Press, 1999)."

```


Db	918	DGAGDXXDDDGKXNADDDTDTGTGTXDDDDKDKMKDMDWDXKGTWGDXTATMAATDWMWWMWMA	855
Oy	238	TTGAGCGCATTAAGGAGGAGAGATTGACGAAACGCTGTCTTACAGTGAAGAAAACAAGT	357
Db	858	DADWTTWMAAADDDWADDDRRMDAMAWMKWDAMWAGARTADRRDWDPRAGKRGARRRDRK	799
Oy	358	TAAACTAATAAGTGA	370
Db	798	RAADKRDAADDD	786
RESULT 8			
LOCUS	CL108838/c	1158 bp	DNA linear GSS 05-JAN-2004
DEFINITION	ISB1-5012_Sp6.1 ISB1 Xenopus tropicalis genomic clone ISB1-5012,		
ACCESSION	CL108838		
VERSION	CL108838.1		
KEYWORDS	GSS.		
SOURCE	Xenopus tropicalis (western clawed frog)		
ORGANISM	Xenopus tropicalis		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;		
	Xenopodine; Xenopus; Silurana.		
	1 (bases 1 to 1158)		
REFERENCE	Kremlicki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,		
AUTHORS	Mardis,E. and Wilson,R.		
	A physical map of the xenopus tropicalis genome		
	Unpublished (2003)		
TITLE	Contact: Richard K Wilson		
JOURNAL	Genome Sequencing Center		
COMMENT	Washington University School of Medicine		
	Email: submissions@watson.wustl.edu		
	Insert Length: 75000 Std Error: 0.00		
	Seq primer: Sp6 ATTAGGTGACACATATG		
	Class: BAC ends		
	High quality sequence start: 36		
	High quality sequence stop: 766.		
FEATURES	Location/Qualifiers		
source	1..1158		
	/organism="Xenopus tropicalis"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:8364"		
	/clone="ISB1-5012"		
	/clone_1ib="ISB1"		
	/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC		
	Library Segment 1"		
ORIGIN			
Query Match	8.6%; Score 43.8; DB 9; Length 1158;		
Best Local Similarity	52.5%; Pred. No.1.6;		
Matches	96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;		
Oy	115	TTGTATTAAGAACCCTCCACACAGTATTAAGTCTATGCAACAAGAAATGCAATACAT	174
Db	781	TGGATTAAGGATCTCACAGAAATGAGGCTCTAAAAAGAAAAAGAAAAA	722
Oy	175	TCTCTAATCATATATTTTCAATTGATAGACCGGTTTTTACTAAGCTCAAAATAG	234
Db	721	GCTGTTGGGTAATTTGTTAAGTTTAAAGGCGCTGTTAAGACCTAAGTTAAATGAAA	662
Oy	235	ATGAACAGATGATGCGTTAGTGACTGTTTATAAAGAAGATAATAAGATCTATCAT	294
Db	661	GTTAAATAATCAATAAGTTATGACATATCTATCCATAGATCAAAATCTTATATTAACCT	602
Oy	295	CAT 297	
Db	601	TAT 599	
RESULT 9			
LOCUS	CG769354	927 bp	DNA linear GSS 29-OCT-2003

DEFINITION	Accession	Version	KeyWords	Source	Organism
TcB41.1_H04_Sp6 Tribolium BAC library Tribolium castaneum genomic, genomic survey sequence.	CG769354	CG769354.1	GI:38022533	GSS.	Tribolium castaneum (red flour beetle)
REFERENCE	Authors	Title	Journal	Comment	
	Savard, J. and Tautz, D.	Tribolium castaneum BAC-ends sequencing project	Unpublished (2003)	Contact: Savard, J. Abteilung für Evolutionsgenetik, AG Tautz Institut für Genetik, Universität zu Köln Weyertal 121, 50931 Köln, Germany Tel.: 49 221 470 6911 Fax: 49 221 470 5975 Email: savard@uni-koeln.de Class: BAC ends.	
FEATURES	Source	Location/Qualifiers			
		1..927			
		/organism="Tribolium castaneum"			
		/mol_type="genomic DNA"			
		/strain="GA-2"			
		/db_xref="taxon:7070"			
		/clone_lib="Tribolium BAC library"			
		/note="vector: pBAC3.6; site_1: EcoRI; site_2: EcoRI; library constructed by Exelixis Inc."			
ORIGIN					
Query Match	8.5%; Score 43.6; DB 9; Length 927;				
Best Local Similarity	53.8%; Pred. No. 1.7;				
Matches	112; Conservative 0; Mismatches 94; Indels 2; Gaps 1;				
Oy	138 TATAAGTCCTATGCAACAAAGAAATGTCATATCTCTAGTCATTATTATTT	197			
Db	26 TATAAGTCAGTTGTAATTTAAACGTGTCATTAATGATTCGTTACATGTCACCTA	85			
Oy	198 CATTAAGATGCGGGTTTCTTACTTCACTCAATTAAGATGAACGATGAATGGTTAGT	257			
Db	86 AATTCGATTTTACACCTGTCACAGAAATACTTTATTTTAAACAAAAGAAACCGATATT	145			
Oy	258 GACGTGTTATTAAGAGAGTAATAA--GATCTATCATCATTTGAGCGAATTAAGCGGAG	315			
Db	146 AAACGGCAATTAACAGATTAATTAACCGGATACATGTTATTACAGTCACGTTGAG	205			
Oy	316 GAGAGATTTCAGCAACAGTGTGCTTACA	343			
Db	206 AAGAGATTACGAAACCGCCGTGTTTCA	233			
RESULT 10					
AV008603	LOCUS	443 bp	mus musculus	EST 25-AUG-1999	
DEFINITION	AV008603	mus musculus 18-day embryo C57BL/6J	mus musculus	cdna	
ACCESSION	AV008603	GI:4785590			
VERSION	AV008603.1				
KEYWORDS	EST.				
SOURCE	mus musculus (house mouse)				
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	Authors	Title	Journal	Comment	
	Carninci, P., Shibata, K., Ozawa, Y., Komu, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,				

[illegible]

RESULT 15			
CR684869/c			
LOCUS	CR684869	1201 bp	mRNA
DEFINITION	Tetrarodon nigroviridis full-length cDNA.		
			linear
			HTC 19-AUG-2004

```

VERSION      CR684869.1  GI:51182776
KEYWORDS     HTC; cDNA; full-length; Tetraodon nigroviridis
SOURCE       Tetraodon nigroviridis
ORGANISM     Tetraodon nigroviridis

```

REFERENCE 1 (bases 1 to 1201)

Submitted (10-4UG-2004) Genoscope - Centre National de Séquençage -
JOURNAL TITLE Direct Submission
AUFORKS Genoscope.

COMMENT The sequences are based on single pass reads.

<http://www.genoscope.cns.fr/tetraodon>.

```
source      1. .1201
/organism="Tetrahodon nigroviridis"
/mol_type="mRNA"
/db_xref="taxon:99883"
/tissue_type="Eyes"
```

ORIGIN

Query Match	8.3%	Score 42.6;	DB 3;	Length 1201;
Best Local Similarity	49.8%	Pred. No. 3.3;		
Matches 108; Conservative	0;	Mismatches 109;	Indels 0;	Gaps 0;

Oy	153	TACCAAGAAATGTCATTCATCTCTTAGTGTCCATTATTTTCATTAGATGCCGGT	212
Db	344	AAAAAAAAAAAAAATATAATTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTATTT	285
Oy	213	TTTTTACTCACTCAATTAGATGAAGAATGATGGTTAGTGACTGTTATATAAGA	272
Db	284	AAAAAAAAAAAAAGTAGATTGGATTAATTTAAAATTTAAAAGGTGTTTTTAAAAAGC	225
Oy	273	AGAGTATTAAGATCTATCATCATTTGAGGCATAAGGAGGAGAGATTCAGCAAACA	332
Db	224	AAAAAAAAAAGAAAAAAGTAAATTTAAAAAAGGGAAAGGAAAGGCGTAAAAAAG	165

```

QY      333  GGTGCTTACAAATGGAAAAACAATTAAACTTAAAGTG 365
          |||  |||  |||  |||  |||  |||  |||  |||
DB      164  TTGTTTAAAAAATTAAAGGAAGGAAAAAGGGG 128

```

Search completed: November 21, 2005, 08:26:16
Job time : 3105.36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2005, 01:59:39 ; Search time 405.736 Seconds
(without alignments)
7470.150 Million cell updates/sec

Title:	US-10-099-663-1_COPY_1115_1626
Perfect score:	512
Sequence:	1 cagtgatccttcattcat.....gtctgctacagacagaag 512

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

1:  N_genseqeq_16Dec04:*
2:  genseqeqn1990s:*
3:  genseqeqn1990s:*
4:  genseqeqn2000s:*
5:  genseqeqn2001as:*
6:  genseqeqn2001bs:*
7:  genseqeqn2002as:*
8:  genseqeqn2002bs:*
9:  genseqeqn2003as:*
10: genseqeqn2003cs:*
11: genseqeqn2003ds:*
12: genseqeqn2004as:*
13: genseqeqn2004bs:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	512	100.0	2381	11	ADL90127	ADL90127 Chicken f
2	326	63.7	336	11	ADL90128	ADL90128 Chicken f
3	42.6	8.3	2000	8	ADA71538	Ada71538 Rice gene
4	42.4	8.3	5908	8	AA545187	AA545187 Chemical1
5	42.4	8.3	5908	6	ABK28232	ABK28232 DNA trans
6	41.8	8.2	10048	6	ABL70313	ABL70313 Chemical1
7	41.8	8.2	10048	6	AA561251	AA561251 Human gen
8	40.6	7.9	7040	4	AA546439	AA546439 Tumour su
9	40.6	7.9	7040	6	ABK33963	ABK33963 Human DNA
10	40.6	7.9	7040	8	ABZ10179	ABZ10179 Haematopo
11	40.6	7.9	7040	8	ABZ10033	ABZ10033 Haematopo
12	40.6	7.9	7040	8	ADA20348	ADA20348 Prostata
13	40.6	7.9	7040	8	ADA84155	ADA84155 Human ren
14	40.6	7.9	7040	10	AD584191	AD584191 Human lym
15	40.6	7.9	7040	10	AD584191	AD584191 Human lym
16	40.6	7.9	14095	6	ABL32477	ABL32477 Human imm
17	40.2	7.9	6065	6	ABK31356	ABK31356 Signal tr
18	40.2	7.9	6065	6	ABL70579	ABL70579 Chemical1
19	40.2	7.9	6065	6	AA561260	AA561260 Human gen
20	40.2	7.9	8365	2	AA220056	AA220056 Plasmodiu

C	21	39.4	7.7	5656	6	ABR28419
C	22	39.4	7.7	110000	10	ACF67367_23
C	23	39.4	7.7	110000	10	ACFG5386_5
C	24	39.2	7.7	9117	6	ABL13069
C	25	39	7.6	2933	4	ABL18616
C	26	39	7.6	5311	6	ABLJ33019
C	27	39	7.6	6240	10	ADDA49062
C	28	39	7.6	22052	4	ABU20542
C	29	39	7.6	26097	4	ABH03696
C	30	39	7.6	36194	4	ABU20582
C	31	39	7.6	59967	4	ABUL15492
C	32	38.4	7.5	32392	6	ABLI56203
C	33	38.4	7.5	50000	6	ABLI55634
C	34	38.2	7.5	6072	6	ABT33031
C	35	38.2	7.5	56737	6	ABS69895
C	36	38.2	7.5	110000	12	A0H63807_4
C	37	38	7.4	6802	6	ABJ33231
C	38	38	7.4	110000	6	ABA92787_1
C	39	37.8	7.4	3322	4	ABI10846
C	40	37.8	7.4	17527	6	ABI33433
C	41	37.8	7.4	17527	6	AAS63333
C	42	37.6	7.3	3296	4	ABU17264
C	43	37.6	7.3	8648	6	ABK31554
C	44	37.6	7.3	8648	6	ABU70573
C	45	37.6	7.3	13573	6	ABJ33868

ALIGNMENTS

RESULT 1						
ID	ADL90127	estandardj; DNA; 2381 BP.				
XX	ADL90127;					
DT	20-MAY-2004	(first entry)				
DE	Chicken intestinal fatty acid binding protein, iFABP, gene, 5' region.					
XX						
KM	Chicken; de; intestinal fatty acid binding protein; iFABP;					
KM	gut specific promoter; transgenic.					
OS	Gallus gallus.					
XX						
PX	US2003177516-AI.					
PD	18-SEP-2003.					
PF	14-MAR-2002; 2002US-00099663.					
PR	14-MAR-2002; 2002US-00099663.					
PA	(HORSEMAN N D.					
PI	(PATENT) PRACT S L.					
PS	Horseman ND, Pract SL;					
DR	WPI; 2003-898653/82.					
PT	New nucleic acid molecule comprising an isolated avian gut-specific gene expression control region, useful for regulating heterologous nucleic acids in transgenic avians, and for generating transgenic birds.					
CC	Claim 1; SEQ ID NO 1; 28bp; English.					
CC	The invention relates to an isolated nucleic acid comprising an isolated avian gut-specific gene expression control region appearing as					
CC	ADL90127/(Chicken intestinal fatty acid binding protein, iFABP, gene, 5'					
CC	region or ADL90128 (chicken iFABP promoter) or its degenerate variant.					
CC	Also included are a recombinant DNA molecule comprising an isolate avian gut-specific gene expression control region operably linked to a nucleic acid insert encoding a polypeptide, an expression vector that integrates					

CC into a host cell (and comprising the isolated avian gut-specific gene
CC expression control region), expressing a heterologous polypeptide in a
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA
CC molecule, and culturing the transfected cell in a medium suitable for
CC expression of a heterologous polypeptide under the control of an avian
CC intestinal fatty acid binding protein (IFABP) or cp35 gene expression
CC control region encoded by the recombinant DNA molecule), a eukaryotic
CC cell transformed with the expression vector (or its progeny, which
CC expresses a heterologous polypeptide) and a transgenic avian having a
CC heterologous polynucleotide sequence comprising the nucleic acid insert.
CC The nucleic acids are useful for regulating heterologous nucleic acids in
CC transgenic avians, as probes in nucleic acid hybridisation assays for
CC detecting the IFABP gene expression control region, and for generating
CC transgenic birds. The present sequence is the Chicken intestinal fatty
CC acid binding protein, IFABP, gene, 5' region.

XX Sequence 2381 BP; 775 A; 428 C; 469 G; 709 T; 0 U; 0 Other;

Query Match 100.0%; Score 512; DB 11; Length 2381;

Best Local Similarity 100.0%; Pred. No. 2.8e-129; Mismatches 0; Gaps 0;

Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGTGATCTTCATCTCATCTTATTTATGTGAGATGATGATTTCCACCAATTAGA 60
DB 1115 CAGTGATCTTCATCTCATCTTATTTATGTGAGATGATGATTTCCACCAATTAGA 1174
QY 61 ATGGAACAAGCAGAGATTTGTTTATCTGTGGTAAATAGTTTCTCCAGTTGAT 120
DB 1175 ATGGAACAAGCAGAGATTTGTTTATCTGTGGTAAATAGTTTCTCCAGTTGAT 1234
QY 121 AAAGACCTCCCAACGATTAAGTCTTATGCAACAAGAAATGTCATATCATCTCTT 180
DB 1235 AAAGACCTCCCAACGATTAAGTCTTATGCAACAAGAAATGTCATATCATCTCTT 1294
QY 181 AGTCTCATTTATTTTTCATTTAGATAGCCGGTTTCTTACCACTCAATTAAGATGAA 240
DB 1295 AGTCTCATTTATTTTTCATTTAGATAGCCGGTTTCTTACCACTCAATTAAGATGAA 1354
QY 241 AGAATAAGTGGTGTAGTGAATGTTTAAAGAGATTAATTAAGATCATCTCATCTT 300
DB 1355 AGAATAAGTGGTGTAGTGAATGTTTAAAGAGATTAATTAAGATCATCTCATCTT 1414
QY 301 AGGCAATTAAGGAGGAGAGATTCAGCAACAAGTGTCTTACCAAGTGAAGAAAGTTAA 360
DB 1415 AGGCAATTAAGGAGGAGAGATTCAGCAACAAGTGTCTTACCAAGTGAAGAAAGTTAA 1474
QY 361 ACTAAAGTGACCCCTCTCTTGAACAAGATCAATGCCAGTTGAGCTTACGAGCCACA 420
DB 1475 ACTAAAGTGACCCCTCTCTTGAACAAGATCAATGCCAGTTGAGCTTACGAGCCACA 1534
QY 421 TCATCATTTATTTGCTTCCGATTAAGCCTGTTCAATTAATTTCTTTGCAAGCTCTGC 480
DB 1535 TCATCATTTATTTGCTTCCGATTAAGCCTGTTCAATTAATTTCTTTGCAAGCTCTGC 1594
QY 481 TACTTACGAGAAGTGTGCTTACAGACAGAAAG 512
DB 1595 TACTTACGAGAAGTGTGCTTACAGACAGAAAG 1626

RESULT 2

ADL90128 standard; DNA; 336 BP.

ADL90128;

20-MAY-2004 (first entry)

Chicken intestinal fatty acid binding protein, IFABP, gene, promoter.

Chicken, ds; intestinal fatty acid binding protein; IFABP;

gut specific promoter; transgenic; promoter.

Gallus gallus.

XX US200317516-A1.
PN 18-SEP-2003.
XX 14-MAR-2002; 2002US-00099663.
PF 14-MAR-2002; 2002US-00099663.
XX 14-MAR-2002; 2002US-00099663.
PR (HORS/) HORSEMAN N D.
XX (PRAT/) PRATT S L.
PA Horseman ND, Pratt SL;
XX WPI; 2003-89653/82.
DR

PT New nucleic acid molecule comprising an isolated avian gut-specific gene
PT expression control region, useful for regulating heterologous nucleic
PT acids in transgenic avians, and for generating transgenic birds.

PS Claim 1; SEQ ID NO 2; 28pp; English.

CC The invention relates to an isolated nucleic acid comprising an isolated
CC avian gut-specific gene expression control region appearing as
CC ADL90127(Chicken intestinal fatty acid binding protein, IFABP, gene, 5'
CC region or ADL90128 (Chicken IFABP promoter) or its degenerate variant.
CC Also included are a recombinant DNA molecule comprising an isolate avian
CC gut-specific gene expression control region operably linked to a nucleic
CC acid insert encoding a polypeptide, an expression vector that integrates
CC into a host cell (and comprising the isolated avian gut-specific gene
CC expression control region), expressing a heterologous polypeptide in a
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA
CC molecule, and culturing the transfected cell in a medium suitable for
CC expression of a heterologous polypeptide under the control of an avian
CC intestinal fatty acid binding protein (IFABP) or cp35 gene expression
CC control region encoded by the recombinant DNA molecule), a eukaryotic
CC cell transformed with the expression vector (or its progeny, which
CC expresses a heterologous polypeptide) and a transgenic avian having a
CC heterologous polynucleotide sequence comprising the nucleic acid insert.
CC The nucleic acids are useful for regulating heterologous nucleic acids in
CC transgenic avians, as probes in nucleic acid hybridisation assays for
CC detecting the IFABP gene expression control region, and for generating
CC transgenic birds. The present sequence is the Chicken intestinal fatty
CC acid binding protein, IFABP, gene, promoter.

XX Sequence 336 BP; 116 A; 62 C; 63 G; 95 T; 0 U; 0 Other;

Query Match 63.7%; Score 326; DB 11; Length 336;

Best Local Similarity 100.0%; Pred. No. 7.9e-79; Mismatches 0; Gaps 0;

Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 ATTATTTATTTTTCATTTAGATAGCCGGTTTTCCTCACTCAATTAAGTGAAGCAATG 246
DB 1 ATTTATTTATTTTTCATTTAGATAGCCGGTTTTCCTCACTCAATTAAGTGAAGCAATG 60
QY 247 AATGGTTAGTACGCTTTTATAAGAGATTAATTAAGTACTATCATCTTTGAGGCA 306
DB 61 AATGGTTAGTACGCTTTTATAAGAGATTAATTAAGTACTATCATCTTTGAGGCA 120
QY 307 TAAGGAGGAGAGATTCAGCAACAAGTGTGCTTCAAGTGAAGAAACAAGTTAACTAA 366
DB 121 TAAGGAGGAGAGATTCAGCAACAAGTGTGCTTCAAGTGAAGAAACAAGTTAACTAA 180
QY 367 GTGACCCCTCTCTTGAACAAGTCAATGCAAGTTGAGCTTTAGCCACCATCATCA 426
DB 181 GTGACCCCTCTCTTGAACAAGTCAATGCAAGTTGAGCTTTAGCCACCATCATCA 240
QY 427 TGTAAATGCTTCCGATTAAGCCTGTTCAATTAATTTCTTTGCAAGCTGTGACTTA 486
DB 241 TGTAAATGCTTCCGATTAAGCCTGTTCAATTAATTTCTTTGCAAGCTGTGACTTA 300
QY 487 CCAGAAGTGTGCTTACAGACAGAAAG 512

Qy	240	CAGAAATGAATGGTGTAGTACTGTTTATTAAGAAGATATAAAGATACTATCATCATTT	299		
Db	776	SRVWKMKKYATRRYYWKKMMATMMMSWRRWKSYSRMWSGWRBRSAMRYCSRMKCAKTKY	711		
Qy	300	GAGCGAATTAAGGAGGAGAGAGATTACGAAACAAGTGTCTTACAACTGGAAAAAAGTTA	355		
Db	716	ASSAATWTRAKRSYSYRRRRWWKRGWTRYRYRWBSCRMTRARMSGRRRWAGASMKSCWM	657		
Qy	360	AACTAAGAAGACCCCCCTGTTGACAAGATCAATGCAAGTTAGACTTTAGACGAC	416		
Db	656	YWRGAKSMWYSKISCSAKCCKTRTWTSSYMTSGMTGMSYTKSMSTSMSTMGKM	600		
RESULT 4					
ID	AA545387/C	standard; DNA, 5908 BP.			
XX	AA545387;				
AC					
XX					
DT	18-DEC-2001	(first entry)			
XX					
DE	Chemically pretreated complementary DNA associated with cell cycle #46.				
XX					
KW	Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;				
KW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;				
KW	graft-versus-host disease; glomerular disease; Lewy body disease; cancer;				
KW	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;				
KW	immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;				
XX	PCR primer.				
XX					
OS	Homo sapiens.				
XX					
PN	WO200168911-A2.				
PD	20-SEP-2001.				
XX					
PF	15-MAR-2001; 2001WO-EP002945.				
XX					
PR	15-MAR-2000; 2000DE-01013847.				
PR	06-APR-2000; 2000DE-01019058.				
PR	07-APR-2000; 2000DE-01019173.				
PR	30-JUN-2000; 2000DE-01032529.				
PR	01-SEP-2000; 2000DE-01043826.				
XX					
PA	(EPIC-) EPIDENOMICS AG.				
XX					
PI	Olek A, Piepenbrock C, Berlin K;				
XX					
DR	WPI; 2001-602751/68.				
XX					
PT	Designing primers and probes for analyzing diseases associated with				
PT	cytosine methylation state e.g. arthritis, cancer, aging,				
PT	arteriosclerosis comprising fragments of chemically modified genes				
XX	associated with cell cycle.				
PS	Claim 1; SEQ ID NO 92; 28pp; English.				
XX					
CC	Sequences AA545387-AA545520 represent chemically pretreated genomic DNA				
CC	molecules associated with the cell cycle and specific PCR primers of the				
CC	invention. The sequences are useful for detecting the methylation state				
CC	of all CpG dinucleotides in a sequence and therefore for analysing				
CC	associated diseases. By analysing cytosine methylations in the pretreated				
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy				
CC	of existing diseases or the predisposition to specific diseases can be				
CC	ascertained. The parameters may be compared to another set of genetic				
CC	and/or epigenetic parameters, the differences serving as basis for				
CC	diagnosis and/or prognosis events which are disadvantageous to patients.				
CC	The sequences of the invention are useful for the diagnosis and therapy				
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,				
CC	aging, glomerular disease, Lewy body disease, arthritis,				
CC	arteriosclerosis, solid tumours and cancers				
XX					
Q0	Sequence 5908 BP; 1706 A; 149 C; 1421 G; 2632 T; 0 U; 0 Other;				

Query Match 8.3%; Score 42.4; DB 4; Length 5908;
Best Local Similarity 54.5%; Pred. No. 0.46; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 138 TATAAGTCCTATGCAACAAAGAAATGCAATACATTCCTTACTCTCATTTATTTT 197
|||||
DB 4703 TATAACTCTTTTATTAATAAACAATATACATATATCTCATTAACCACTTTTAA 4644
|||||

QY 198 CATTAGATAGCCGCTTTTCTACACCTCAATTAAGTAAAGACAGATGAATGGTTAGT 257
|||||
DB 4643 TATTATATTTTATTTTACACTACCTTCTCAAAATTTAAATTAATCTATACCTTAAT 4584
|||||

QY 258 GACTGTTTATAAAGAGATTAATTAAGATCTATCA 293
|||||
DB 4583 TCCCTTTATCCAAATTAATTAATCTATATACGACA 4548
|||||

RESULT 5
ABK28232/c
ID ABK28232 standard; DNA; 5908 BP.

AC ABK28232;
XX
DT 23-APR-2002 (first entry)
XX

DE DNA transcription associated complementary genomic DNA #53.
XX

KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX

OS Unidentified.
XX

PN WO200192565-A2.
XX

PD 06-DEC-2001.
XX

PF 06-APR-2001; 2001WO-EP003973.
XX

PR 06-APR-2000; 2000DE-01019058.
XX
PR 07-APR-2000; 2000DE-01019173.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX

PA (EPIG-) EPIGENOMICS AG.
XX

PI Olek A, Piepenbrock C, Berlin K;
XX

DR MPI; 2002-090046/12.
XX

PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX

PS Claim 1; SEQ ID NO 106; 32pp; English.
XX

CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for

CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status) e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office

SEQ Sequence 5908 BP; 1706 A; 149 C; 1421 G; 2632 T; 0 U; 0 Other;

Query Match 8.3%; Score 42.4; DB 6; Length 5908;
Best Local Similarity 54.5%; Pred. No. 0.46;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 138 TATAAGTCCTATGCAACAAAGAAATGCAATACATTCCTTACTCTCATTTATTTT 197
|||||
DB 4703 TATAACTCTTTTATTAATAAACAATATACATATATCTCATTAACCACTTTTAA 4644
|||||

QY 198 CATTAGATAGCCGCTTTTCTACACCTCAATTAAGTAAAGACAGATGAATGGTTAGT 257
|||||
DB 4643 TATTATATTTTATTTTACACTACCTTCTCAAAATTTAAATTAATCTATACCTTAAT 4584
|||||

QY 258 GACTGTTTATAAAGAGATTAATTAAGATCTATCA 293
|||||
DB 4583 TCCCTTTATCCAAATTAATTAATCTATATACGACA 4548
|||||

RESULT 6
ABL70313/c
ID ABL70313 standard; DNA; 10048 BP.

AC ABL70313;
XX

DT 01-JUN-2002 (first entry)
XX

DE Chemically treated cell signalling DNA sequence#102.
XX

DE Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytosine; ds.
XX

OS Unidentified.
XX

PN WO200202807-A2.
XX

PD 10-JAN-2002.
XX

PF 29-JUN-2001; 2001WO-EP007471.
XX

PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX

PA (EPIG-) EPIGENOMICS AG.
XX

PI Olek A, Piepenbrock C, Berlin K;
XX

DR MPI; 2002-154758/20.
XX

PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling.
XX

PS Claim 1; SEQ ID NO 203; 24pp + Sequence Listing; English.
XX

CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the

CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL7011-ABL7062 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office

XX Sequence 10048 BP; 2873 A; 57 C; 1888 G; 5225 T; 0 U; 5 Other;

Query Match 8.2%; Score 41.8; DB 6; Length 10048;

Best Local Similarity 46.4%; Pred. No. 0.81;

Matches 136; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

XX 100 ATACGTTTCTCCAGTTGTAAGACCTCCACAGTAATGAAGTCTATGCAACAAG 159

DB 5471 ATATTATATCTCTAACTAAATTCCTCTACAAAAAAATATATATCTATAC 5412

QY 160 AAAATGCAATACATCTCTAGTCTCATATATTTTCTATGATAGCCGTTTTCAC 219

DB 5411 TCCAAATCCAAAAATCTATTTTAAACCTATTAATTCCTTACAAATACCTTCTAC 5352

QY 220 TACAACTCAATAGATGAAGATGAATGAGTGTAGTACTGTTATTAAGAAGATA 279

DB 5351 AATAACTAATATAAAAACACATTAATTCCTTAATCTAATATCAAAATACAAAAAAC 5292

QY 280 TAAAGTACTATCATCATTTAGGCAATTAAGGAGGAGAGATTGACGAAACAGTGTCT 339

DB 5291 TATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5232

QY 340 TCAAGTGAAGAAACAGTTAAATTAAGTACCCCTCTGACAAAGTCA 392

DB 5231 AAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 5179

RESULT 7

AA561251/c

ID AA561251 standard; DNA; 10048 BP.

XX AA561251;

DT 29-JAN-2002 (first entry)

DE Human gene regulation-associated gene oligonucleotide #206.

XX Human; Gene regulation-associated gene; severe combined immunodeficiency;

KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;

KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;

KW renal disease; Preclampsia; cardiac allograft vascular disease;

KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;

KW immunostimulant; cardiant; antiinflammatory; coagulant; antiaesthetic;

KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

XX Homo sapiens.

OS MO200177375-A2.

XX 18-OCT-2001.

PF 06-APR-2001; 2001MO-EP003968.

XX 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-0103529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-017470/02.

DR New nucleic acid sequences from chemically modified genes associated with

PT gene regulation, useful for analyzing cytosine methylations for diagnosis

XX and therapy of diseases e.g. severe combined immunodeficiency disease.

PS Claim 1; SEQ ID NO 211; 26bp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least

CC 18 bases of a chemically pretreated gene associated with gene regulation

CC selected from 43 known genes (or complementary sequences). The chemical

CC pretreatment converts cytosine bases unmethylated at the 5-position to

CC uracil or another base with hybridisation behaviour dissimilar to

CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,

CC oligomers (or sets/arrays) and method are useful in the diagnosis of

CC diseases (or predisposition to diseases) associated with gene regulation

CC and in therapy of such diseases, by enabling analysis of the cytosine

CC methylation patterns of such genes. Kits are provided. They are

CC especially useful in diagnosis and therapy of e.g. severe combined

CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours

CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen

CC syndrome, renal disease, preclampsia, graft versus-host disease. The

CC present sequence is a sequence included in the sequence data for this

CC specification and is associated with the human gene regulation-associated

CC gene. Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WPI at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 10048 BP; 2873 A; 57 C; 1888 G; 5225 T; 0 U; 5 Other;

Query Match 8.2%; Score 41.8; DB 6; Length 10048;

Best Local Similarity 46.4%; Pred. No. 0.81;

Matches 136; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

XX 100 ATACGTTTCTCCAGTTGTAAGACCTCCACAGTAATGAAGTCTATGCAACAAG 159

DB 5471 ATATTATATCTCTAACTAAATTCCTCTACAAAAAAATATATATCTATAC 5412

QY 160 AAAATGCAATACATCTCTAGTCTCATATATTTTCTATGATAGCCGTTTTCAC 219

DB 5411 TCCAAATCCAAAAATCTATTTTAAACCTATTAATTCCTTACAAATACCTTCTAC 5352

QY 220 TACAACTCAATAGATGAAGATGAATGAGTGTAGTACTGTTATTAAGAAGATA 279

DB 5351 AATAACTAATATAAAAACACATTAATTCCTTAATCTAATATCAAAATACAAAAAAC 5292

QY 280 TAAAGTACTATCATCATTTAGGCAATTAAGGAGGAGAGATTGACGAAACAGTGTCT 339

DB 5291 TATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5232

QY 340 TCAAGTGAAGAAACAGTTAAATTAAGTACCCCTCTGACAAAGTCA 392

DB 5231 AAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 5179

XX Homo sapiens.

OS AA546439/c

ID AA546439 standard; DNA; 7040 BP.

XX AA546439;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #161.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;

KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

XX cytosine methylation; ds.

XX Homo sapiens.

XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 161; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pre-treated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumor suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (S8) and sequences
CC complementary to (S8). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7040 BP; 2018 A; 184 C; 1566 G; 3272 T; 0 U; 0 Other;
XX
XX Query Match 7.9%; Score 40.6; DB 4; Length 7040;
XX Best Local Similarity 50.8%; Pred. No. 1.5; Indels 0; Gaps 0;
XX Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
XX
QY 101 TACGTTTCTCCAGTTGATTAAGACCCCTCCACAGTAAAGTCTATGCAACAAGA 160
DB 3586 TATATTTTAAATAATCAATTAATAATTCCTTTCTTAACACCTTAATTAAT 3527
QY 161 AATATGCAATCAATTCCTTAGCTCATTTATTTTCATATAGATGCGGTTTACT 220
DB 3526 ATATTAACCACTTTTAATTAATACAAATTCATTTTAAATAATTAATTAATTA 3467
QY 221 ACAACTCAATTAAGATGACAGATGAATGAGTGAAGTCTTTTAAAGAAAGTAA 280
DB 3466 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3407
QY 281 AAGATATCTAT 291
DB 3406 AATATTTCTAT 3396

RESULT 9
ABK33963/c
ID ABK33963 standard; DNA; 7040 BP.

XX
AC ABK33963;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human DNA for staging of Astrocytomas #24.
XX
KW Human; ds; astrocytoma; cytosine; staging; cytosine methylation; CpG;
KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
KW matrix assisted laser desorption/ionization mass spectrometry.
XX
OS Homo sapiens.
XX
PN WO200202808-A2.
XX
PD 10-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007538.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-171649/22.
XX
XX Novel chemically modified genomic DNA sequences, useful in the
PT characterization, classification, differentiation, grading, staging,
PT treatment and/or diagnosis of astrocytomas or predisposition to
PT astrocytomas.
XX
PS Claim 1; SEQ ID NO 47; 37pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence (I) of at
CC least 18 bases in length of a segment of chemically pre-treated genomic
CC DNA which has any one of the sequences of (ABK33919-ABK34032) or its
CC complement. Also included are an oligonucleotide or peptide nucleic acid
CC (or set thereof) of at least 9 nucleotides which hybridises to (I),
CC primers for (I), probes for detecting cytosine methylation or single-
CC nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide
CC nucleic acids for analysing diseases associated with the methylation
CC states of the CpG dinucleotides of (I). The array is useful for
CC determining genetic and/or epigenetic parameters, classification,
CC differentiation, grading, staging, treatment and/or diagnosis of
CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine
CC methylations, involves obtaining a biological sample containing genomic
CC DNA, extracting the genomic DNA, converting cytosine bases which are
CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or
CC another base which is dissimilar to cytosine in terms of hybridisation
CC behaviour, by chemical treatment and amplifying chemically pre-treated
CC genomic DNA fragments using the array and a polymerase, where the
CC amplification carries a detectable label. The method further involves
CC identifying methylation status of one or more cytosine positions, and
CC analysing methylation status of the cytosine positions by reference to
CC one or more data sets. The genomic DNA is chemically treated by using a
CC bisulphite, hydrogen sulphite or disulphite. The amplification step
CC amplifies DNA which is of particular interest in astrocytoma or brain
CC tissue, based on the specific genomic methylation status of brain
CC tissues, as opposed to background DNA. The amplification carries a
CC fluorescent label or radionuclide. Optionally, the labels of the
CC amplification are detachable molecule fragments having a typical mass
CC which are detected in a mass spectrometer. The fragments of chemically
CC pre-treated genomic DNA to be amplified, have a single positive or
CC negative charge for a better detectability in the mass spectrometer.
CC Preferably, the amplification or fragments of the amplification are
CC detected by matrix assisted laser desorption/ionization mass spectrometry
CC (MALDI) or using electron spray mass spectrometry (ESI). The present
CC sequence is one of the chemically pre-treated reference DNA samples of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX	SQ	Sequence	7040 BP; 2018 A; 184 C; 1566 G; 3272 T; 0 U; 0 Other;
XX		Query Match	7.9%; Score 40.6; DB 6; Length 7040;
XX		Best Local Similarity	50.8%; Pred. No. 1.5;
XX		Matches	97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
OY		101	TACGTTTCTCCAGTGATGTATAAGACCCTCCCACCAAGTAAAGTCCTATGCACAAGA 160
DB		3586	TATATTAAAAAATCAAATATATAAAATTCCTCTTCTAAAACAATTATATTAAT 3527
OY		161	AAAGTCATATCATTCTCTGATGCATTATATTTTCATTAGATAGCGGGTTTTACT 220
DB		3526	ATATATTAACCATTTTAAATTAACAATTCATTTTAAAAATATTAACCTTAAAAA 3467
OY		221	ACAACCTCAATTAAGTGAAGCAATGAATGGGTTAGTACGTGTTTATTAAGAGATAT 280
DB		3466	ATTAATTAATAAACCTTAACCATTTAGCAAAAACAATTTAAATCTATTAACCTTAAT 3407
OY		281	AAAGTACTAT 291
DB		3406	AAAAATTTCTAT 3396
CC		RESULT 10	
CC		ABZ10179/c	
XX	ID	ABZ10179	standard; DNA; 7040 BP.
XX	XX	ABZ10179;	
XX	DT	16-JAN-2003	(first entry)
DE	XX	Haematopoietic cell proliferation disorder related DNA sequence #319.	
KW	XX	Human; haematopoietic cell proliferation disorder; cytostatic;	
KM	XX	gene therapy: lymphocytic leukaemia; acute myelogenous leukaemia;	
KW	XX	cytosine methylation scale; gene; ds.	
OS	XX	Homo sapiens.	
PN	XX	WO20027272-A2.	
XX	PD	03-OCT-2002.	
XX	PF	26-MAR-2002; 2002WO-EP003401.	
XX	PR	26-MAR-2001; 2001US-0278333P.	
PA	XX	(EPIG-) EPIGENOMICS AG.	
PB	XX	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller U;	
PI	PI	Olek A, Pispembrock C, Adorian P, Grabs G, Ieasche R, Leu B;	
PI	PI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelot C;	
PI	PI	Schweppe I, Ziebarth H;	
DR	XX	WPI; 2003-018942/01.	
PT	XX	Detecting and differentiating between hematopoietic cell proliferative	
PT	XX	disorders, comprises contacting a target nucleic acid with a reagent that	
PT	XX	distinguishes between methylated and non-methylated CpG dinucleotides.	
PS	XX	Claim 28; SEQ ID NO 319; 117bp; English.	
CC	XX	The present invention describes a method for detecting and	
CC	XX	differentiating between haematopoietic cell proliferative disorders	
CC	XX	associated with at least 1 gene and/or their regulatory regions in a	
CC	XX	subject. The method comprises contacting a target nucleic acid in a	
CC	XX	biological sample obtained from the subject with at least 1 reagent,	
CC	XX	which distinguishes between methylated and non-methylated CpG	
CC	XX	dinucleotides within the target nucleic acid. AB209861 to AB21118	
CC	XX	represent specifically claimed nucleotide sequences from the present	
CC	XX	invention. Oligonucleotides from the present invention can be used for	
CC	XX	differentiating between healthy haematopoietic cells and proliferative	

CC	disorder haematopoietic cells; for differentiating between acute
CC	lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC	determining the cytosine methylation state and/or single nucleotide
CC	polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC	related sequences and their complements; and as primers for the
CC	amplification of haematopoietic cell proliferation disorder related DNA
CC	sequences. The nucleotide sequences from the present invention can also
CC	be used for detecting a predisposition to, differentiation between
CC	subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC	haematopoietic cell proliferative disorders. The present method enables a
CC	highly specific classification of haematopoietic cell proliferative
CC	disorders allowing for improved and informed treatment of patients
SX	
SQ	Sequence 7040 BP; 2018 A; 0 C; 1566 G; 3456 T; 0 U; 0 Other;
Query Match	7.9%; Score 40.6; DB 8; Length 7040;
Best Local Similarity	50.8%; Pred. No. 1.5;
Matches	97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
OY	101 TACGTTTCTCCAGCTGTATTAAGAAGCCTCCCAACGATTAAGTCCTATGAACAAGA 160
Dd	3586 TATATTTTAAATAAATCAATTAATAAATTCCTTCCTAAAACATTATATTAAT 3527
OY	161 AAAGTCATATCATTCCTTAGTCTCATTTATTTTCATTAGATAGCGGTTTTACT 220
Dd	3526 ATATATTAACCACTTTAATTAATTAACATTCATTTTAAAAATATTAACCTTAAAAA 3467
OY	221 ACAACCTAAATTAAGATGACAGATGAATGGGTTAGACGCTTTATTAAGAAGAGTAAT 280
Dd	3466 ATTAATTAATTAATTAACATTAACCATTACAAAAAACATTTAAATCTATAAAAA 3407
OY	281 AAAGATCTACTAT 291
Dd	3406 AAAATTCTCAT 3396
Db	
RESULT 11	
ABZ10033/C	
ID	ABZ10033 standard; DNA; 7040 BP.
XX	
AC	ABZ10033;
DT	16-JAN-2003 (first entry)
DE	Haematopoietic cell proliferation disorder related DNA sequence #173.
KM	Human; haematopoietic cell proliferation disorder; cytostatic;
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
RN	Cytosine methylation state; gene; da.
OS	Homo sapiens.
XX	
PN	WO200277272-A2.
PD	03-OCT-2002.
PF	26-MAR-2002; 2002WO-BP003401.
XX	
PR	26-MAR-2001; 2001US-0278333P.
PA	(EPIC-) EPIDEMIOLOGICS AG.
P1	Berlin K., Braun A., Distler J., Guetig D., Howe A., Mueller J;
P1	Olek A., Piepenbrock C., Adorjan P., Grabs G., Lesche R., Leu B,
P1	Lewin A., Lippecker B., Mater S., Model F., Mueller V., Otto T., Pelet C,
P1	Schwope I., Ziesarth H;
XX	
DR	WPI; 2003-018942/01.
PT	Detecting and differentiating between hematopoietic cell proliferative
PT	disorders, comprises contacting a target nucleic acid with a reagent that
XT	distinguishes between methylated and non-methylated CpG dinucleotides.

PA (EPIC-) EPIGENOMICS AG.
XX
XX
PI Distler U, Model F, Adorjan P;
XX
DR WPI; 2003-167536/16.
XX
PT Determining genetic and/or epigenetic parameters, useful for the
PT classification, differentiation and/or diagnosis of prostate tumors or a
PT predispotion to prostate cancer, comprises analyzing cytosine
PT methylation.
XX
XX
PS Claim 28; Page 68-72; 376pp; English.
XX
XX The invention relates to a method of determining genetic and/or
CC epigenetic parameters for the classification, differentiation and/or
CC diagnosis of prostate tumors or the predispotion to prostate cancer,
CC by analysing cytosine methylation in a sample of genomic DNA. The method
CC comprises chemically treating unmethylated cytosine bases at the 5-
CC position to uracil or another base, which is dissimilar to cytosine in
CC terms of hybridization behaviour; followed by amplifying at least one
CC fragment of the chemically pre-treated genomic DNA using sets of primer
CC oligonucleotides and a polymerase. The oligomers or probes derived from
CC them are useful for detecting the methylation state of all CpG
CC dinucleotides and/or single nucleotide polymorphisms (SNPs) in a
CC chemically pre-treated genomic DNA. They are all useful for treating
CC prostate carcinoma. This sequence represents a fragment of genomic DNA

Sequence 7040 BP; 2018 A; 184 C; 1566 G; 3372 T; 0 U; 0 Other;

	Query Match	7.9%	Score 40.6;	DB 8,	Length 7040;	
	Best Local Similarity	50.8%;	Pred. No.1.5;			
	Matches	97;	Conservative	0;	Mismatches	94;
					Indels	0;
					Gaps	0
QY	101	TACGTTTCTCCAGTGTATAAGCCTCCCAACAGATAAAGTCATTGCACAAGA	160			
DB	3586	TATATTTAAAAAATCAAATATAATAATTTCCCTTCTCTAAAACCTTATATATTAAT	352			
QY	161	AAATGCAATCATCTCTTAGTGTCATTATTA TTTCTATTAGATAGCGGGTTTTACT	220			
DB	3526	ATATATAACCACATTTTATATAATAACAATTCATTTTAAAAATATATTAACCTTAAAAA	346			
QY	221	ACAACCAATTAAGTGAACAGAAATGAATGGGTAGTACTGTTATATAAGAGATAAT	280			
DB	3466	ATAAATTATAAACATTAACCATTTTACAAAAACAATTTAAATCTATATAAACCTTAAAT	340			

Db 3406 AAAATTCTAT 3396

ADA84155/C
ID ADA84155 standard; DNA; 7040 BP.
XX

XX	20-NOV-2003	(first entry)
DT		
XX		
XX	Human renal/prostate carcinoma associated DNA SEQ ID NO:13	
DE		
XX	da; renal cancer; prostate cancer; cytosine methylation;	
KW	single nucleotide polymorphism; histological; cytological.	
XX		

XX
PN
XX

WO2002103041-A2.

FD	27-DEC-2002.
XX	
PF	14-JUN-2002; 2002WO-EP006603.
XX	

Query Match	Best local Similarity	7.9%	Score 40.6;	DB 10;	Length 7040;
Matches	97;	Conservative	0;	Mismatches	94;
				Indels	0;
				Gaps	0;
Oy	101	TACGTTTTTCCAGTTGATATAAGAGCCCTCCACAGATATAAGTCTATGCAACAAGA	160		
Db	3586	TATATTTTAAAAAATCAAAATATATTAATTTCCCTCTCTTAAACACTTATATATTAAT	3527		
Oy	161	AAATGCAATACATCTCTTAGTCTCATTTATTTTCATTAGATAGACGGTTTTTACT	220		
Db	3526	ATATATTAACCACTTTTAAATTAATTAACAACTCTATTTTAAATATATTAATCTTAAAAA	3467		
Oy	221	ACAACTCAATTAAGATGAACGAATGAATGGTTAGAGCTGTTTATTAAGGAAGATAT	280		
Db	3466	ATTAATTAATTAACATTAACCACTTTACAAAAACAACTTTATTAATAACCTTAAT	3407		
Oy	281	AAAGATACATAT	291		
Db	3406	AAAAATTTCTAT	3396		

ID ADE84115 standard; DNA; 7040 BP.
XX
AC ADE84115;
XX
XX 29-JAN-2004 (first entry)
XX
DE Human lymphoid cell proliferative disorder gene derived DNA #51.
XX
XX de; lymphoid cell proliferative disorder; methylation;
KM methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KM diffuse large B-cell lymphoma; mantle cell lymphoma;
KM chronic lymphocytic leukemia; small lymphocytic lymphoma;
KM follicular lymphoma; diagnosis; prognosis.
XX
OS Homo sapiens.
XX
PN WO2003044226-A2.
XX
PD 30-MAY-2003.
XX
PF 25-NOV-2002; 2002WO-EP013265.
XX
PR 23-NOV-2001; 2001DE-01057491.
XX
PR 28-DEC-2001; 2001DE-01064501.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
PI Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;
XX
DR WPI; 2003-457621/43.
XX
XX
PT Detecting and differentiating between lymphoid cell proliferative
PT disorders comprises contacting a target nucleic acid with at least one
PT reagent that distinguishes between methylated and non-methylated CpG
PT dinucleotides.
XX
PS Claim 26; SEQ ID NO 111; 448bp; English.
XX
XX
CC The invention relates to a method of detecting and differentiating
CC between lymphoid cell proliferative disorders associated with at least
CC one gene and/or their regulatory regions in a subject by contacting a
CC target nucleic acid in a biological sample obtained from the subject with
CC at least one reagent or series of reagents that distinguish between
CC methylated and non-methylated CpG dinucleotides within the target nucleic
CC acid. The genes and/or their regulatory regions are preferably selected
CC from MDRI, CSNK2B, EGR4, AR, CDK4, RB2, CDC25A, GRB2, MYO1, CDH3,
CC MYCL1, EIK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, FOS,
CC GSTP1, HIC-1, MGMT, MLH1, MDS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKN1C,
CC GSK3beta, ESRI, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
CC acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC of the genes are useful for detecting the methylation state of all the
CC CpG dinucleotides within one or more the sequences, or their complements,
CC for determining the cytosine methylation state and or single nucleotide
CC polymorphisms (SNPs), and for differentiating at least two of the medical
CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,
CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
CC lymphoma. They are also useful for detecting of a predisposition to,
CC differentiation between subclases, diagnosis, prognosis, treating and/or
CC monitoring of lymphoid cell proliferative disorder. This sequence
CC represents a nucleic acid of a pretreated genomic DNA derived from the
CC above mentioned genes.
XX
SQ Sequence 7040 BP; 2018 A; 184 C; 1566 G; 3272 T; 0 U; 0 Other;

Query Match 7.9%; Score 40.6; DB 10; Length 7040;
Best Local Similarity 50.8%; Pred. No. 1.5;
Matches 97; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 101 TACGTTTCCGAGTATATAAGACCCGCAAGTATATAAGTCTATGCAAGAAAGA 160
DB 3586 TATATTTTAAATAAATAAATAATTCCTTTCTTAATAACACTTATATATTAATTT 3527
QY 161 AAATGCAATACATTCCTTAGTCTCATTTATTTTCATTAGATAGCCGGTTTAACT 220

DB 3526 ATATATTAACCAATTTTATATATAACAAATCTATTTTAAATATATTAACCTTAAAAAA 3467
QY 221 ACAACTCAATTAAGATGAACGAATGAATGGGTAGTGAATGCTTTATATAAGAAAGTAAAT 280
DB 3466 ATTAATTTAATAAACAATAACCAATTTTAAATCTATTAATAAACCTTAAAT 3407
QY 281 AAAGATTAAT 291
DB 3406 AAAATTTCTAT 3396

Search completed: November 21, 2005, 05:53:43
Job time : 410.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2005, 05:12:21 ; Search time 82.8113 Seconds
(without alignments)
6639.036 Million cell updates/sec

Title: US-10-099-663-2

Perfect score: 336
Sequence: 1 attatatttttcatgata.....agacagaagaatgcattra 336

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgm2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgm2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgm2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgm2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgm2_6/prodata/1/ina/6C_COMB.seq:*
6: /cgm2_6/prodata/1/ina/backfillseq1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.2	11.1	1141	US-09-806-708B-22	Sequence 22, Appl
2	36.8	11.0	43117	US-09-949-016-17589	Sequence 17589, A
3	36.2	10.8	1587	US-09-710-279-1929	Sequence 1929, Ap
4	36.2	10.8	1878	US-09-134-001C-1667	Sequence 1667, Ap
5	36.2	10.8	3000	US-09-710-279-3907	Sequence 3907, Ap
6	36.2	10.8	3368	US-09-710-279-4169	Sequence 4169, Ap
7	36	10.7	168394	US-09-949-016-13002	Sequence 13002, A
8	35.8	10.7	7218	US-08-232-463-14	Sequence 14, Appl
9	35.6	10.6	601	US-09-949-016-44395	Sequence 44395, A
10	35.6	10.6	601	US-09-949-016-58898	Sequence 58898, A
11	35.6	10.6	139562	US-09-949-016-13451	Sequence 13451, A
12	35.6	10.6	256171	US-09-949-016-12822	Sequence 12822, A
13	35.6	10.6	256176	US-09-949-016-15524	Sequence 15524, A
14	35.2	10.5	601	US-09-949-016-134974	Sequence 134974, A
15	35.2	10.5	663	US-08-998-416-187	Sequence 187, App
16	35.2	10.5	666	US-08-998-416-779	Sequence 779, App
17	35.2	10.5	719	US-08-998-416-1138	Sequence 1138, Ap
18	35.2	10.5	856	US-08-998-416-289	Sequence 289, App
19	35	10.4	399	US-09-621-976-8976	Sequence 8976, Ap
20	34.8	10.4	78125	US-09-949-016-16006	Sequence 16006, A
21	34.6	10.3	601	US-09-949-016-173718	Sequence 173718, A
22	34.6	10.3	601	US-09-949-016-173765	Sequence 173765, A
23	34.6	10.3	81819	US-09-949-016-16661	Sequence 16661, A
24	34.6	10.3	81819	US-09-949-016-16661	Sequence 16661, A
25	34.4	10.2	1137	US-09-107-532A-2846	Sequence 2846, Ap
26	34.4	10.2	1664976	US-08-916-421B-1	Sequence 1, Appl
27	34.4	10.2	1664976	US-09-692-570-1	Sequence 1, Appl

28	34.2	10.2	98844	US-09-791-211-10	Sequence 10, Appl
29	34.2	10.2	143776	US-09-949-001-29	Sequence 29, Appl
30	34.2	10.2	144034	US-09-949-001-35	Sequence 35, Appl
31	33.6	10.0	2119	US-09-240-639-7	Sequence 7, Appl
32	33.6	10.0	2119	US-09-908-510A-7	Sequence 7, Appl
33	33.6	10.0	2119	US-09-905-744B-7	Sequence 7, Appl
34	33.6	10.0	2119	US-10-107-660-7	Sequence 7, Appl
35	33.6	10.0	2119	US-10-107-576-7	Sequence 7, Appl
36	33.6	10.0	2119	US-09-905-732B-7	Sequence 7, Appl
37	33.6	10.0	2119	US-09-905-743B-7	Sequence 7, Appl
38	33.4	9.9	31385	US-09-949-016-13310	Sequence 13310, A
39	33.4	9.9	73788	US-09-949-016-12358	Sequence 12358, A
40	33	9.8	2642	US-08-178-242-4	Sequence 4, Appl
41	33	9.8	2642	US-08-955-091-4	Sequence 4, Appl
42	33	9.8	2642	US-09-225-510-4	Sequence 4, Appl
43	33	9.8	2870	US-08-178-242-14	Sequence 14, Appl
44	33	9.8	2870	US-08-955-091-14	Sequence 14, Appl
45	33	9.8	2870	US-09-225-510-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806, 708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147, 133
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.c., I.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match      11.1%; Score 37.2; DB 4; Length 1141;
Best Local Similarity 10.7%; Pred. No. 0.24; Mismatches 135; Indels 0; Gaps 0;
Matches 31; Conservative 123;

QY 1 ATTATATTCTTCTAGATAGCCGGTTTCTTCTCACTCAATTAAGTGAACAGATG 60
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 285 AACMCRABAYGWNABVNSTCTTWKSKTTRYSWMANNCAGDANKDKHKKWSBAAG 344

QY 61 AATGGCTTAGTACTGTCTTATTAAGAAGATATAAGTACTATCATATTGAGCA 120
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 345 VVNNNNNNNNNTYKXABHBAWVWHSAMKXHANAAHYSRKKWTBYRKTNNNGT 404

QY 121 TAGGAGAGGAGAGATTCAGCAACAGCTGCTTCAAGTGAAGAAAAGTTAACTAA 180
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 405 TTMKRWAMWYKMDMDWBGTYNNNNNGRTYGYTKKKWYTKKANNCKRWADHKT 464

QY 181 GTGACCCCTCTTGCAGCAAGATCAATGACAGTTGAGTTCAGCAGCATCATCA 240
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 465 CTNNNTTWMKMTWNNNCYKSMTGKSHRBAAYVYTWMMWRRAHANNNNMDYMWKA 524

QY 241 TGTAAATTCCTTCCGATTAACCTGTTGATTAATTCCTTTGCAAGC 289
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 525 CTWYKBYCCKWNNWYAAWYTKSSWNYTSWNYTSWYRKTNNSWRWSRSDTRSM 573

RESULT 2
```



```
US-09-949-016-17589/c
; Sequence 17589, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17589
; LENGTH: 43117
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17589

Query Match
Best Local Similarity 52.6%; Pred. No. 1.3;
Matches 80; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 30 TTACTCAACTCAATTAAGTGAACAGATGAATGGCTTGTGACTGTTTATTAAGAGA 89
DB 20807 TGAACCAATATATATGATGATGAATAATATGAGGATTAATTAAGTATTAATGCA 20748
QY 90 GTAATAAGTACTATCATCATTTGAGGCAATTAAGGAGGAGAGATTGACAAACAGTG 149
DB 20747 CCTTAAGAGCAATTAATGTAAGTGAATATGCTCTGCGGAAATCTGAAGACAGA 20688
QY 150 TGCTTCAAGTGAACAAAGTTAACTAAG 181
DB 20687 GTTTAAAGATGACAAACCAATTAGATGAG 20656

RESULT 3
US-09-710-279-1929
; Sequence 1929, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1929
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1929

Query Match
Best Local Similarity 10.8%; Score 36.2; DB 4; Length 1587;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAAAGGTGCTTACAGTGAAGAAACAGTTAACTAAAGTACCCCTCTTGACAA 200
DB 203 CAAACATGACTTTAAAGTTGAACCTTATTTTACAAAGTACCTCAAGCCCATCTA 262
QY 201 GATCAATGCCAGTTGAGCTTAGCCAGCCACATCATCATGTAAATGCTTCTCTGATA 260
```

```
DB 263 GAAAAATTAACAAATTAATCTCAAGCATATATCATCATCAATTAATCACTTGAA 322
QY 261 AGCTGTTCATTAATTCCTCTTGCA 285
DB 323 ATAAATACAAATTAATCTTGCA 347

RESULT 4
US-09-134-001C-1667
; Sequence 1667, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1667
; LENGTH: 1878
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1667

Query Match
Best Local Similarity 10.8%; Score 36.2; DB 3; Length 1878;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 141 CAAAGGTGCTTACAGTGAAGAAACAGTTAACTAAAGTACCCCTCTTGACAA 200
DB 494 CAGACGATGACTTTAAAGTTGAACCTTATTTTCAAGAGTTACCTCAAGCCCATCTA 553
QY 201 GATCAATGCCAGTGTGAGCTTTACCGCCACATCATCATTAATGCTTCTGATA 260
DB 554 GAAAAATTAACAAATTAATCTCAAGCATATATGATATCATCAATTAATCACTTGAA 613
QY 261 AGCTGTTCATTAATTCCTCTTGCA 285
DB 614 ATAAATACAAATTAATCTTGCA 638

RESULT 5
US-09-710-279-3907
; Sequence 3907, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3907
; LENGTH: 3000
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3907

Query Match
Best Local Similarity 10.8%; Score 36.2; DB 4; Length 3000;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
```


QY 141 CAACAGCTGCTTACAGTGAAGAAACAGTTAAAGTGAACCCCTCTTGACAA 200
| | | | |
DB 129 CAGACGATGCTTTAAAGTTGAACCTTATTTTACAGGTTACCTCAAGCCCATACTA 188
| | | | |
QY 201 GATCAATGACAGTGTAGCTTAGCCAGCCATCATCTGTAATGCTTCTTGATA 260
| | | | |
DB 189 GAAAAATTAACCAATTAATCTCAAGACATATAGTATATCAATTCATCACTTCTGAAA 248
| | | | |
QY 261 AGCCTGTTCAATTAATCTCTTTGCA 285
| | | | |
DB 249 ATAAATACAAACATTAATCTGTGCA 273
| | | | |
RESULT 6
US-09-710-279-4169/c
; Sequence 4169, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4169
; LENGTH: 3368
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-4169
Query Match 10.8%; Score 36.2; DB 4; Length 3368;
Best Local Similarity 53.1%; Pred. No. 0.72;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 141 CAACAGCTGCTTACAGTGAAGAAACAGTTAAAGTGAACCCCTCTTGACAA 200
| | | | |
DB 1038 CAGACGATGCTTTAAAGTTGAACCTTATTTTACAGGTTACCTCAAGCCCATACTA 979
| | | | |
QY 201 GATCAATGACAGTGTAGCTTAGCCAGCCATCATCTGTAATGCTTCTTGATA 260
| | | | |
DB 978 GAAAAATTAACCAATTAATCTCAAGACATATAGTATATCAATTCATCACTTCTGAAA 919
| | | | |
QY 261 AGCCTGTTCAATTAATCTCTTTGCA 285
| | | | |
DB 918 ATAAATACAAACATTAATCTGTGCA 894
| | | | |
RESULT 7
US-09-949-016-13002/c
; Sequence 13002, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13002
; LENGTH: 168394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(168394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13002
Query Match 10.7%; Score 36; DB 4; Length 168394;
Best Local Similarity 58.3%; Pred. No. 3.7;
Matches 63; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 27 TTTTACTACAACTCAATTAAGATGAACAGAAATGAGTTAGTACTGTTTATTAAGA 86
| | | | |
DB 124517 TTTTGCCTCAATTAAGATGAATGATCAATGAATTAAGAAATAGTGTCTTATAGAGT 124458
| | | | |
QY 87 AGAGTAAATTAAGTACTATCATCTTTTGAGGCAATTAAGGAGGAGAG 134
| | | | |
DB 124457 TTATTTATTTTAAATCTCTTAACAATAGTATGAAGGAGTCAAG 124410
| | | | |
RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29, 768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpC-Fls
US-08-232-463-14
Query Match 10.7%; Score 35.8; DB 1; Length 7218;

```
Best Local Similarity 7.4%; Pred. No. 1.3;
Matches 13; Conservative 100; Mismatches 62; Indels 0; Gaps 0;

QY 10 TTCATTAGATGACCGGTTTTCATCAACTCAATAAGATGAACGAATGAGGTTA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1459 TTTAAAGATGAGAAATTTGGTACRRTTTRRTTTRRTTTRRTTTRRTTTR 1400
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 GTGACGTTTATAAAGATATAAGATATCATCATCTTTGAGCAATGAGGAG 129
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1399 RRRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 GAGAGATTCAGCAACAGTGTCTTCAAGTGGAATACTAACTAAAGTGA 184
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1339 RRRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTRRTTTR 1285
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-949-016-44395/C
; Sequence 44395, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; LENGTH: 601
; SEQ ID NO 44395
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-44395

Query Match 10.6%; Score 35.6; DB 4; Length 601;
Best Local Similarity 57.4%; Pred. No. 0.57;
Matches 62; Conservative 1; Mismatches 45; Indels 0; Gaps 0;

QY 27 TTTTCTACACTCAATAAGATGAACGAATGAGGTTGCTGACTGTTTATAAGA 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 398 TTTTGACCTCAATAGATGAATCAATGAATAAGAAATGCTCTTTATAGAGT 339
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 AGAGTATAAAGATCATCATCATCTTTGAGCAATGAGGAGGAG 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 TTTATTATTATTATTCTTCAATGAGTAAAGGAGTCAGAG 291
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-949-016-58898
; Sequence 58898, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 58898
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-58898

Query Match 10.6%; Score 35.6; DB 4; Length 601;
Best Local Similarity 52.7%; Pred. No. 0.57;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 43 AATAAGATGAAAGAAATGAGTGTGCTTTTATAAAGAGTAAATGAATAC 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38132 ATTTCATGAGATGAGATGATGATGCTTTTGAATATTTATACCAATGAAAAGT 38191
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 TATCATCTTTGAGGCAATTAAGGAGGAGATTCAGCAACAGTGTCTTACAAGTGG 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38192 TATATGATGCCAGAAATGATGACCAAGAGCAAGAAAGAGAGGACAAAGTTC 38251
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 AAAACAGTTAAACTAAAGTGACCCC 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38252 AAAAGCTGTAGAGGCCAGAGAGACC 38277
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-09-949-016-13451
; Sequence 13451, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13451
; LENGTH: 139562
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.(139562)
; LOCATION: (1)..(139562)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13451

Query Match 10.6%; Score 35.6; DB 4; Length 139562;
Best Local Similarity 52.7%; Pred. No. 4.6;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 43 AATAAGATGAAAGAAATGAGTGTGCTTTTATAAAGAGTAAATGAATAC 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38132 ATTTCATGAGATGAGATGATGATGCTTTTGAATATTTATACCAATGAAAAGT 38191
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 TATCATCTTTGAGGCAATTAAGGAGGAGATTCAGCAACAGTGTCTTACAAGTGG 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38192 TATATGATGCCAGAAATGATGACCAAGAGCAAGAAAGAGAGGACAAAGTTC 38251
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 AAAACAGTTAAACTAAAGTGACCCC 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 38252 AAAAGCTGTAGAGGCCAGAGAGACC 38277
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-949-016-12822/C
; Sequence 12822, Application US/09949016
; Patent No. 6812339
```

```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 12822
; LENGTH: 256171
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(256171)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12822

Query Match      10.6%; Score 35.6; DB 4; Length 256171;
Best Local Similarity 49.0%; Pred. No. 5.7;
Matches 95; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 4 ATTATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 63
DB 88458 ACTGATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 88399
QY 64 GGGTTAGTGAAGCTGTTTATAAAGAGTAATAAGATATCATCTATTTGAGGCAATTA 123
DB 88398 ATGTAAATATTTGTGATCATATTTAAAAACCAAAACCTATTTTATGATCATGTAA 88339
QY 124 GGGAGGAGAGATTACGCAAAACAGTGTGCTTACAGTGGAAAAACAGTTAACTAAAGTG 183
DB 88338 ATATATCATATCCATGTTTAAAGTTTGAAATTCACAGAAACATGATGAAAAAATTA 88279
QY 184 ACCCCCTCCTTGA 197
DB 88278 AACTACCATTTAA 88265

RESULT 13
US-09-949-016-15524/C
; Sequence 15524, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 15524
; LENGTH: 256176
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(256176)
; OTHER INFORMATION: n = A,T,C or G
```

```

US-09-949-016-15524

Query Match      10.6%; Score 35.6; DB 4; Length 256176;
Best Local Similarity 49.0%; Pred. No. 5.7;
Matches 95; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 4 ATTATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 63
DB 88458 ACTGATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 88399
QY 64 GGGTTAGTGAAGCTGTTTATAAAGAGTAATAAGATATCATCTATTTGAGGCAATTA 123
DB 88398 ATGTAAATATTTGTGATCATATTTAAAAACCAAAACCTATTTTATGATCATGTAA 88339
QY 124 GGGAGGAGAGATTACGCAAAACAGTGTGCTTACAGTGGAAAAACAGTTAACTAAAGTG 183
DB 88338 ATATATCATATCCATGTTTAAAGTTTGAAATTCACAGAAACATGATGAAAAAATTA 88279
QY 184 ACCCCCTCCTTGA 197
DB 88278 AACTACCATTTAA 88265

RESULT 14
US-09-949-016-134974
; Sequence 134974, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ. ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 134974
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-134974

Query Match      10.5%; Score 35.2; DB 4; Length 601;
Best Local Similarity 48.5%; Pred. No. 0.75;
Matches 94; Conservative 1; Mismatches 99; Indels 0; Gaps 0;

QY 4 ATTATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 63
DB 232 ACTGATTTTCATTGATAGCGGGTTTTTACTACAACTCAATATAGATGAACAGATTAAT 291
QY 64 GGGTTAGTGAAGCTGTTTATAAAGAGTAATAAGATATCATCTATTTGAGGCAATTA 123
DB 292 ATGTAAATATTTGTGATCATATTTAAAAACCAAAACCTATTTTATGATCATGTAA 351
QY 124 GGGAGGAGAGATTACGCAAAACAGTGTGCTTACAGTGGAAAAACAGTTAACTAAAGTG 183
DB 352 ATATATCATATCCATGTTTAAAGTTTGAAATTCACAGAAACATGATGAAAAAATTA 411
QY 184 ACCCCCTCCTTGA 197
DB 412 AACTACCATTTAA 425

RESULT 15
US-08-998-416-187/C
; Sequence 187, Application US/08998416
; Patent No. 6239264
```

GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-Dec-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-Dec-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 187:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1074UP
US-08-998-416-187

Query Match 10.5%; Score 35.2; DB 3; Length 663;
Best local Similarity 58.7%; Pred. No. 0.78; Mismatches 43; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 1 ATTATTATTTCTATTAGATGCCGTTTCTACTCAACTCAATATAGATGACAGATG 60
DB 567 ATTATTAACCTTATTAGTTAAACCATTTATTATGATCATTAATATAATAAGAAATA 508
QY 61 AATGGTTAGTACTGCTTTATTAAGAAGACTAATAAGATACTA 104
DB 507 CATTAATGATATATAGTTATTATAGAACCAATGAAGATACTA 464

Search completed: November 21, 2005, 08:29:58
Job time : 86.8113 secs


```
Db 1 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 60
Qy 61 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 120
Db 61 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 120
Qy 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 180
Db 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 180
Qy 181 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 240
Db 181 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 240
Qy 241 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 300
Db 241 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 300
Qy 301 CCAGAACTGTGCTTACAGACAGAAAGATGCAATTTA 336
Db 301 CCAGAACTGTGCTTACAGACAGAAAGATGCAATTTA 336

RESULT 2
US-10-099-663-1
; Sequence 1, Application US/10099663
; Publication No. US2003017516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian GUT-Specific Promoters
; FILE REFERENCE: A181
; CURRENT APPLICATION NUMBER: US/10/099,663
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2381
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(1626)
; NAME/KEY: exon
; LOCATION: (1627)..(1693)
; NAME/KEY: Intron
; LOCATION: (1694)..(2322)
; NAME/KEY: exon
; LOCATION: (2323)..(2381)
US-10-099-663-1

Query Match 100.0%; Score 336; DB 17; Length 2381;
Best Local Similarity 100.0%; Pred. No. 4,1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 60
Db 1301 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 1360
Qy 61 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 120
Db 1361 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 1420
Qy 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 180
Db 1421 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 1480
Qy 181 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 240
Db 1481 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 1540
Qy 241 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 300
Db 1541 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 1600
```

```
Qy 301 CCAGAACTGTGCTTACAGACAGAAAGATGCAATTTA 336
Db 1601 CCAGAACTGTGCTTACAGACAGAAAGATGCAATTTA 1636

RESULT 3
US-09-925-065A-572321/C
; Sequence 572321, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 572321
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-572321

Query Match 12.7%; Score 42.8; DB 13; Length 558;
Best Local Similarity 53.6%; Pred. No. 0.11;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 60
Db 346 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 287
Qy 61 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 120
Db 286 TAAGAGTAAGTCATGATTAATTAAGATTAAGAACTTAATTAAGATTAAGCA 227
Qy 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAA 166
Db 226 TAAGAACTTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGCA 181

RESULT 4
US-09-925-065A-875384/C
; Sequence 875384, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875384
; LENGTH: 558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875384/C

Query Match 100.0%; Score 336; DB 17; Length 2381;
Best Local Similarity 100.0%; Pred. No. 4,1e-85;
Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 60
Db 1301 ATTATATTTCATTAGATAGCCGGTTTCTACTCAACTCAATTAAGATGACAGAAATG 1360
Qy 61 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 120
Db 1361 AATGGTTAGTACCTTTTATAAGAGTAATAAGTACTATCATATTTGAGGCA 1420
Qy 121 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 180
Db 1421 TAAGGAGGAGAGATTTCAGCAACAGTGTGCTTACAACTGGAACAAAGTTAACTAA 1480
Qy 181 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 240
Db 1481 GTGACCCCTCTCTTGCACAAAGTCAATGCGACAGTTGAGCTTACGACCCATCATCA 1540
Qy 241 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 300
Db 1541 TGTAAATGCTTCCGATTAAGCTGTTCATTAATCTCTTTGCAAAAGCTGTGCTACTTA 1600
```

```

; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875384
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875384
```

```

Query Match
Best Local Similarity 12.7%; Score 42.8; DB 13; Length 679;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```

QY 1 ATTATATTTTCATTAGATAGCCGGTTTCTTACCACTCAATATAGATGACAGATG 60
DB 515 ATTAACTTAAACAATTAGAGCTTTATATCTCTTAAACAAGACATTTAGAGACT 456
QY 61 AATGGGTAGTGACTGTTTAAAGAAGAGTAAATAGATCATCATCTTTAGGCAA 120
DB 455 TAAGAGTAAGTCCATGTATATAGATTAAGATTAAGAACACTATATATAGATTAAGCAA 396
QY 121 TAAGGAGGAGAGGATTCAGCAACAGTGTCTTACAGTGGAATA 166
DB 395 TAAGAACACTTAATCTTGAATAATTAATTAATAAATGTACAAA 350
```

RESULT 5

```

US-09-925-065A-875385/C
; Sequence 875385, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108627.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875385
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875385
```

```

Query Match
Best Local Similarity 12.7%; Score 42.8; DB 13; Length 679;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```

QY 1 ATTATATTTTCATTAGATAGCCGGTTTCTTACCACTCAATATAGATGACAGATG 60
DB 515 ATTAACTTAAACAATTAGAGCTTTATATCTCTTAAACAAGACATTTAGAGACT 456
QY 61 AATGGGTAGTGACTGTTTAAAGAAGAGTAAATAGATCATCATCTTTAGGCAA 120
DB 455 TAAGAGTAAGTCCATGTATATAGATTAAGATTAAGAACACTATATATAGATTAAGCAA 396
QY 121 TAAGGAGGAGAGGATTCAGCAACAGTGTCTTACAGTGGAATA 166
DB 395 TAAGAACACTTAATCTTGAATAATTAATTAATAAATGTACAAA 350
```

RESULT 6

```

US-09-925-065A-875386/C
```

```

; Sequence 875386, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108627.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 875386
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-875386
```

```

Query Match
Best Local Similarity 12.7%; Score 42.8; DB 13; Length 679;
Matches 89; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
```

```

QY 1 ATTATATTTTCATTAGATAGCCGGTTTCTTACCACTCAATATAGATGACAGATG 60
DB 515 ATTAACTTAAACAATTAGAGCTTTATATCTCTTAAACAAGACATTTAGAGACT 456
QY 61 AATGGGTAGTGACTGTTTAAAGAAGAGTAAATAGATCATCATCTTTAGGCAA 120
DB 455 TAAGAGTAAGTCCATGTATATAGATTAAGATTAAGAACACTATATATAGATTAAGCAA 396
QY 121 TAAGGAGGAGAGGATTCAGCAACAGTGTCTTACAGTGGAATA 166
DB 395 TAAGAACACTTAATCTTGAATAATTAATTAATAAATGTACAAA 350
```

RESULT 7

```

US-10-719-993-6867
; Sequence 6867, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Cargill, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6867
; LENGTH: 116327
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(116327)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-10-719-993-6867
```

```

Query Match
Best Local Similarity 12.0%; Score 40.4; DB 21; Length 116327;
Matches 77; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
```

```

QY 197 ACAAGATCAATGCCAGTGTAGCTTTAGCCGACGACATCATCATTAATGCTTCT 256
```

```
Db 70844 AAAAGATGATGCCCTCACTAGGTTGACACAGTATTTAGCATCATAGATTACCT 70903
QY 257 GATTAAGCCGTTCATTAATCTCTTGGCAAGCTGCTACTACAGAAAGTCTGCTAC 316
Db 70904 GGTAACTTTTAAATAATTCATTCCCAAGTCTTACTTAATAAAAAAGATCAATTTT 70963
QY 317 AGACAGAAAGATGTCATT 334
Db 70964 AGAAAGAAAAAGATATT 70981

RESULT 8
US-11-097-143-25165
; Sequence 25165, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25165
; LENGTH: 2933
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-25165

Query Match 11.6%; Score 39; DB 26; Length 2933;
Best Local Similarity 52.1%; Pred. No. 3;
Matches 87; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 73 ACTGTTTAAAGACAGTAATTAAGATCTATCATCATTTGAGCAATAAGGAGGAG 132
Db 2766 AATTTAAAGACAGATCTTACTTAAGCAATAATAGGATTAATGATGAGAAAGAGAGAT 2825
QY 133 AGATTGAGCAAAAGTGTGTTTCAAGTGAAAAAAGTTAACTTAAGTACCCCTC 192
Db 2826 AATTTAAAGCAAAATTTCTTATTCATGTAAGAAAGAGAACCGAAATTTGCATTATC 2885
QY 193 CTGACAAAGATCAATCCACAGTTGAGCTTTAGCCAGCAATCATC 239
Db 2886 CGTGGCCCACTGAATGCAATCACTTAGTGACACCAACCATACCAACC 2932

RESULT 9
US-09-925-065A-115086
; Sequence 115086, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
```

```
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115086
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-115086

Query Match 11.1%; Score 37.4; DB 13; Length 593;
Best Local Similarity 49.8%; Pred. No. 4.1;
Matches 123; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 52 AACAGATGAATGGGTTAGTGCCTTTTAAAGAAGTATTAAGATCATCATCAT 111
Db 290 AACAGAAAAAGATATTTTAAAGTATTAATCTAGAAAGAGAGATTAATCATCAT 349
QY 112 TTGAGGCAATAGGAGGAGAGATTTCAGCAACAGTGTCTTACAGTGAACAACT 171
Db 350 TTTTCAATATGATATAGCAATTTAATCTAGAAACAGAGACTGAATGAATTTTAT 409
QY 172 TAACTTAAAGTACCCCTCTTGAACAAGATCAATGCCAGTTGAGCTTTAGCCAGCC 231
Db 410 AAACGTGATTTTAAAGTCACTAAATGCAATGCAATGCCAA--GACACATAGGAATCA 466
QY 223 AATCATCATGTAATTTGTTCTTCGATTAAGCTGTTCATTAATCTCTTGCAGAACTC 291
Db 467 AAGCTTTTGGACACTAAATTAAGGACATTTCTAATTTAAATTTCTTTACAAAGTA 526
QY 292 TGTCTACT 298
Db 527 AGCAGCT 533

RESULT 10
US-09-925-065A-115087
; Sequence 115087, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115087
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
```


US-09-925-065A-115087

Query Match 11.1%; Score 37.4; DB 13; Length 593;
Best Local Similarity 49.8%; Pred. No. 4.1;
Matches 123; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 52 AACAGATGATGAGTGTAGTCTGTTTATTAAGAGATTAATTAAGTACTATCAT 111
DB 290 AACAGAAAAAGATATTTTAAAGATTAATCTAGAAAAGAGATTAATCATCAT 349
QY 112 TTGAGCAATTAAGGAGGAGAGATTGAGCAAAAGTGTGCTTACAGTGAAGAAAGT 171
DB 350 TTTTGCAATGATATGATGCAATTTAATGAAAAACAGAGACTTAATGAAATTTAT 409
QY 172 TAACTTAAGTGAACCCCTCTTGTGACAGATTAATGCCAGTTGACCTTTCAGACC 231
DB 410 AAAACGATTTTAAAGTTGATTAAGTGAAGTGCACAAAGCCAAA--GACACATAGGAATCA 466
QY 232 ACATCATCATGTAATGCTTCTGCTGATAGCCGTGCTCAATTAATGCTTTGCAAGCTC 291
DB 467 ACAGCTTTTGGACACTTAATTAAGGAGATTTCTTAATTAATTTCTTTACAAAAGTA 526
QY 292 TGCTACT 298
DB 527 AGCAGCT 533

RESULT 11

US-09-925-065A-115089
; Sequence 115089, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115089
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-115089

Query Match 11.1%; Score 37.4; DB 13; Length 593;
Best Local Similarity 49.8%; Pred. No. 4.1;
Matches 123; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 52 AACAGATGATGAGTGTAGTCTGTTTATTAAGAGATTAATTAAGTACTATCAT 111
DB 290 AACAGAAAAAGATATTTTAAAGATTAATCTAGAAAAGAGATTAATCATCAT 349
QY 112 TTGAGCAATTAAGGAGGAGAGATTGAGCAAAAGTGTGCTTACAGTGAAGAAAGT 171
DB 350 TTTTGCAATGATATGATGCAATTTAATGAAAAACAGAGACTGAATGAAATTTAT 409
QY 172 TAACTTAAGTGAACCCCTCTTGTGACAGATTAATGCCAGTTGACCTTTCAGACC 231
DB 410 AAAACGATTTTAAAGTTGATTAAGTGAAGTGCACAAAGCCAAA--GACACATAGGAATCA 466
QY 232 ACATCATCATGTAATGCTTCTGCTGATAGCCGTGCTCAATTAATGCTTTGCAAGCTC 291

DB 467 ACAGCTTTTGGACACTTAATTAAGGAGATTTCTATTAATTAATTTCTTTACAAAAGTA 526

QY 292 TGCTACT 298
DB 527 AGCAGCT 533

RESULT 12

US-11-097-143-7087
; Sequence 7087, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7087
; LENGTH: 2927
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-7087

Query Match 11.1%; Score 37.2; DB 26; Length 2927;
Best Local Similarity 51.9%; Pred. No. 9.7;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 73 ACTGTTTATTAAGAGATTAATTAAGATTAATCATCATTTGAGCAATTAAGGAGAG 132
DB 2766 ACATTAATGACAGATCTTACTTAACGAATTAATGATTAATGATGACGAAGAGAGAT 2825
QY 133 AGATTGAGCAACAGAGTGTCTTACAGTGAAGAAAGATTAAGTGAACCCCTC 192
DB 2826 ATATTACAGAAATTTCTTATGATGATGAAGCAAGTGAACCAATTTGATTCATC 2885
QY 193 CTGACAAATCATGATGCCAGCTTGAAGCTTGAAGCCAGCCACA 234
DB 2886 CGTGCCCACTGATGCAATCATTAAGTGAACCAACCCATA 2927

RESULT 13

US-09-910-185-11/C
; Sequence 11, Application US/09910185
; Publication No. US20030083279A1
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-3 EXPRESSION
; FILE REFERENCE: RTS-0258
; CURRENT APPLICATION NUMBER: US/09/910,185
; CURRENT FILING DATE: 2001-07-18

NUMBER OF SEQ ID NOS: 90
SEQ ID NO 11
LENGTH: 123526
TYPE: DNA
ORGANISM: Homo sapiens
US-09-910-185-11

Query Match 11.1%; Score 37.2; DB 10; Length 123526;
Best Local Similarity 53.4%; Pred. No. 53;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 12 CATTAGATGACCGGTTTCTTACTACATCAATTAAGTGAACAGAAATGAGTGTACT 71
DB 107264 CATTCCCAATGCTTTTAAATGCTACAGAAATGCAAGAAAGGGGTGCT 107205
QY 72 GACTGTTTAAAGAGAGTAATTAAGTACTTCTATTGAGGCATTAAGGAGGGA 131
DB 107204 GTGTACCCATTAAATTAATGAGATGAAGAAAGAAAGGCAAGGAGGGA 107145
QY 132 GAGATTCAGCAACAGTGTCTTACA 157
DB 107144 GGGAGAAACCGAAAGCATCATGCA 107119

RESULT 14
US-09-925-065A-115088
Sequence 115088, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 115088
LENGTH: 593
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-115088

Query Match 11.0%; Score 37; DB 13; Length 593;
Best Local Similarity 49.4%; Pred. No. 5.3;
Matches 122; Conservative 1; Mismatches 121; Indels 3; Gaps 1;

QY 52 AACAGATGATGGCTTAGTACTGTTTAAAGAGTAATTAAGTACTATCAT 111
DB 290 AACAGAAAAAGATATTTTAAAGTAACTAGTAAAGAAAGATTAATCATCAT 349
QY 112 TTGAGCAATATAGGAGGAGATTCAGCAACAGTGTCTTACAAGTGAACAAG 171
DB 350 TTTTGCAATGATATAGCAATTTTACTAGAAACACAGAGACTGAATGAATTTAT 409
QY 172 TAACTAAAGTGAACCCCTCTCTGACAGATCATGCAAGTGAAGTGTAGCCAGCC 231
DB 410 AAAAGTATTTTAAAGTGAATGAATGCAAGTGAAGTGAAGTGAAGTGAAGTGA 466
QY 232 ACATCATGATGTAATGCTTCTGATTAAGCTGTTTCAATTAATTTCTTTGAAAGCT 291
DB 467 ACAGCTTTTGCAAGTAAATTAAGGCAAGTCTTAATTTAAATTTCTTTTACAAAAGTA 526

QY 292 TGCTACT 298
DB 527 AGCACT 533

RESULT 15
US-09-925-065A-701240/c
Sequence 701240, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 701240
LENGTH: 1846
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-701240

Query Match 11.0%; Score 37; DB 13; Length 1846;
Best Local Similarity 52.2%; Pred. No. 8.9;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 16 AGATAGCCCGTTTTTCTACCACTCAATTAAGTGAACAGAAATGAGTGTAGTACT 75
DB 1271 AGATAGCTTCAATTTTAAATCTCTAAMCCATTAAGTGAAGTGAAGTGAAGTGA 1212
QY 76 GTTATTAAGAGAGTAATTAAGATCTATCTATTTGAGGCAATTAAGGAGGAGAGA 135
DB 1211 GAATTAACATAAAAACTATGATTCATGACTTTTCAAAAATTTCTGATTAAGAGA 1152
QY 136 TTCAGCAACAGTGTCTTACAAGTGAACAAGTT 172
DB 1151 AGTAAATATGCTGAAACCAATGAAGAAATTT 1115

Search completed: November 21, 2005, 11:55:02
Job time: 541.283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:48:44 ; Search time 9972 Seconds
(without alignments)
11569.582 Million cell updates/sec

Title: US-10-099-663-1
Perfect score: 2381
Sequence: 1 agctccgcgcgcgaagaag.....atctgaagctactatctcag 2381

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_ph:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_sg:*
11: gb_st:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	736.8	30.9	2847	5	AY254202 Gallus ga
2	85	3.6	923	5	BX932340 Gallus ga
3	81	3.4	899	5	BX930518 Gallus ga
4	78.8	3.3	896	5	BX933597 Gallus ga
5	64.2	2.7	1141	6	AX083744 Sequence
6	62	2.6	3013	10	RATFABP
7	62	2.6	230897	2	AC130496
8	57.8	2.4	1141	6	AX083744
9	56.6	2.4	5039	10	MUSFABP
10	51.8	2.2	234081	3	PFMALP2
11	50	2.1	564	10	RATFABP
12	49.8	2.1	619	5	XELIFABP
13	49.8	2.1	5204	6	AX771590
14	49.8	2.1	5204	9	HUMFABP
15	49.8	2.1	200000	2	AC008077
16	49.2	2.1	2000	6	AX655393
17	49	2.1	136061	2	AC117821
18	49	2.1	202804	10	AC021630
19	49	2.1	249943	3	AE014823

20	48.4	2.0	564	10	RATFABP	K01180 Rat intesti
21	48.4	2.0	832	6	AR415176	AR415176 Sequence
22	48.4	2.0	832	6	AX972010	AX972010 Sequence
23	48.4	2.0	832	6	BD110729	BD110729 EST and e
24	48.4	2.0	18876	3	CERF3258	Z72509 Caenorhabdi
25	48.4	2.0	178045	2	CERF62472	CR762472 Danio rer
26	48.4	2.0	197110	9	AC104306	AC104306 Homo sapi
27	48.2	2.0	132254	3	AC116330	AC116330 Dictyoste
28	48.2	2.0	132254	3	AC116330	AC116330 Dictyoste
29	48	2.0	191840	9	AC092656	AC092656 Homo sapi
30	47.8	2.0	110000	3	AC116984	Continuation (2 of
31	47.8	2.0	137174	2	BX276186	BX276186 Danio rer
32	47.8	2.0	162810	2	CR392028	CR392028 Danio rer
33	47.6	2.0	53302	3	AC116961	AC116961 Dictyoste
34	47.6	2.0	143364	2	BX324220	BX324220 Danio rer
35	47.6	2.0	174176	5	AL772340	AL772340 Zebrafish
36	47.6	2.0	177623	2	CR394571	CR394571 Danio rer
37	47.4	2.0	110000	2	PFMAL13	Continuation (10 o
38	47.4	2.0	318221	2	PFMAL13P3	AL049184 Plasmodiu
39	47.2	2.0	165797	2	CR847851	CR847851 Danio rer
40	47	2.0	2636	6	CQ593303	CQ593303 Sequence
41	47	2.0	81181	2	AC019668	AC019668 Drosophi
42	47	2.0	176969	3	AC010707	AC010707 Drosophi
43	47	2.0	181815	5	BX248120	BX248120 Zebrafish
44	47	2.0	184977	5	BX649334	BX649334 Zebrafish
45	47	2.0	192540	3	AC010846	AC010846 Drosophi

ALIGNMENTS

RESULT 1
AY254202 2847 bp DNA linear VRT 22-APR-2003
LOCUS AY254202
DEFINITION Gallus gallus intestinal fatty acid-binding protein gene, complete cds.

ACCESSION AY254202
VERSION AY254202.1 GI:30060211
KEYWORDS
SOURCE
ORGANISM Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE
Wang, Q., Li, H., Wang, Y. and Zhao, J.
1 (bases 1 to 2847)
TITLE Cloning and characterization of chicken I-FABP gene
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2847)
Wang, Q., Li, H., Wang, Y. and Zhao, J.
Direct Submission
TITLE Submitted (11-MAR-2003) Animal Science & Technology, Northeast Agricultural University, 59 Mucai Street, Harbin, Heilongjiang 150030, China

FEATURES
source
location/Qualifiers
1..2847
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/tissue_type="blood"
/note="Breed: Arbor Acres broiler"
join(1..21..87,720..892,1465..1572,2416..2466)
/product="intestinal fatty acid-binding protein"
join(21..87,720..892,1465..1572,2416..2466)
/note="FABP"

CDS
mRNA
ORIGIN
1..2847
/product="intestinal fatty acid-binding protein"
/protein_id="AAPI3101.1"
/db_xref="GI:30060212"
/translation="MAFGTGWKIEKNYERKMEAMGVNVMKRLGADNLTLLTQQD
GNKFLVKSNNPRTIDIEFTGVSEYSLADGTELSGSWNLEGNLTGVGTFPKDKNGKV
LTAYEITGSELTCTYVVEGVAKRIFPKE"

Query Match 30.9%; Score 736.8; DB 5; Length 2847;
 Best Local Similarity 98.1%; Pred. No. 4e-146;
 Matches 757; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

QY 1613 CTACAGACAGAAAGATGCGATTAAACGCTACTTGGAAAAATAGAGAAATGAGAACTATG 1672
 DB CGAGAGACAGAAAGATGCGATTAAACGCTACTTGGAAAAATAGAGAAATGAGAACTATG 66

QY 1673 AAAAAATTCATGAGAGCAATGGGTAGCCCTTACTTTTGAATGCCCTTCTAAAAGCAGAT 1732
 DB 67 AAAAAATTCATGAGAGCAATGGGTAGCCCTTACTTTTGAATGCCCTTCTAAAAGCAGAT 126

QY 1733 ACCACAGCGCGGGAATACAAACTTAAAGCTTCAATGCTTACCATCTGGCTAACCTGTC 1792
 DB 127 ACCACTACGCGCGGGAATACAAACTTAAAGCTTCAATGCTTACCATCTGGCTAACCTGTC 186

QY 1793 TTTGTGTCTGCTATTTTCCCTTGCACATTCGCCCTGCACCTTATTTTGAAGAACTCTAT 1852
 DB 187 TTTGTGTCTGCTATTTTCCCTTGCACATTCGCCCTGCACCTTATTTTGAAGAACTCTAT 246

QY 1853 AGAGGGGAATACAGAGAGAAAAAATTCTGATTTTATTTGCTTCCGTAATCTTATGC 1912
 DB 247 AGAGGGGAATACAGAGAGAAAAAATTCTGATTTTATTTGCAATGCGATAATCTTATGC 306

QY 1913 ATTAGCTAATTCAGTAGAGGCAATCCAGCAGAAATTTAAATAGAAATTAATGTAAGA 1972
 DB 307 ATTAGCTAATTCAGTAGAGGCAATCCAGCAGAAATTTAAATAGAAATTAATGTAAGA 366

QY 1973 ATATATTTTGTATAGACTGTTTGAATAATTAACAAGAGGGAATTTGCTGCTCCACT 2032
 DB 367 ATATATTTTGTATAGACTGTTTGAATAATTAACAAGAGGGAATTTGCTGCTCCACT 426

QY 2033 TTTGCGAACAACATGATTTT--GAGTCAATTTTAAACAGCTAGTCTTACTTTAAGCTT 2089
 DB 427 TTTGCGAACAACATGATTTTGAATTTTAAACAGCTAGTCTTACTTTAAGCTT 486

QY 2090 GTACAAATGCGCTGTAATATGATGTAAACATACTATCCTAGTTGGATAGTATGTA 2149
 DB 487 GTACAAATGCGCTGTAATATGATGTAAACATACTATCCTAGTTGGATAGTATGTA 546

QY 2150 TTACAGGCTGAACACTGCTCAGTGAAGGTGAGAGAGTAAGACTGTAAGTCAAAAT 2209
 DB 547 TTACAGGCTGAACACTGCTCAGTGAAGGTGAGAGAGTAAGACTGTAAGTCAAAAT 606

QY 2210 CTGGGCTAAGCTCCCTCACTACAGAAAAAGTCAATATAAATGCAACATGATGTTCT 2269
 DB 607 CTGGGCTAAGCTCCCTCACTACAGAAAAAGTCAATATAAATGCAACATGATGTTCT 666

QY 2270 ATTTGTGTTTCTCTGCTGATGATTAATGATTAATTTTATTTTATTTTATTTTATTTT 2329
 DB 667 ATTTGTGTTTCTCTGCTGATGATTAATGATTAATTTTATTTTATTTTATTTTATTTT 726

QY 2330 TGTGATGAAAAAGAACTTGAAGCCACAGATAATCTGAAGCTCACTATTCAAG 2381
 DB 727 TGTGATGAAAAAGAACTTGAAGCCACAGATAATCTGAAGCTCACTATTCAAG 778

RESULT 2
 BX932340 923 bp mRNA linear VRT 30-MAR-2004
 LOCUS BX932340
 DEFINITION Gallus gallus finished cDNA, clone CHEST55022.
 ACCESSION BX932340
 VERSION BX932340.2 GI:46017395
 KEYWORDS
 SOURCE
 ORGANISM
 Gallus gallus (chicken)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus
 1 (bases 1 to 923)
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,
 Croning, M.D.R., Davies, R.M., Francis, R.M., Grahame, D.V.,
 Hubbard, S.J., Humphrey, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
 Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,
 Tickle, C. and Wilson, S.A.
 Direct Submission
 Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: chickens@hinxton.ac.uk
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
 sequencing project.
 This sequence is from the
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
 from a library constructed by Elizabeth Bosch. cDNA was prepared
 from RNA extracted from heart, normalised, and poly A-tailed.
 EcoRI-NotI cut cDNA was then ligated into the vector. Vector:
 pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host: Escherichia
 coli DH10B.
 Location/Qualifiers
 1..923
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hixex"
 /db_xref="taxon:9031"
 /clone="CHEST355022"
 /clone_1ib="CSBOCHN70"
 /dev_stage="stage 36"

ORIGIN

Query Match 3.6%; Score 85; DB 5; Length 923;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1610 TGCCACAGACAGAAAGATGCGATTAAACGCTACTTGGAAAAATAGAGAAATGAGAACT 1669
 DB 1 TGCCACAGACAGAAAGATGCGATTAAACGCTACTTGGAAAAATAGAGAAATGAGAACT 60

QY 1670 ATGAAAAATTCATGAGAGCAATGGG 1694
 DB 61 ATGAAAAATTCATGAGAGCAATGGG 85

RESULT 3
 BX930518 899 bp mRNA linear VRT 02-FEB-2004
 LOCUS BX930518
 DEFINITION Gallus gallus finished cDNA, clone CHEST640B17.
 ACCESSION BX930518
 VERSION BX930518.1 GI:41631046
 KEYWORDS
 SOURCE
 ORGANISM
 Gallus gallus (chicken)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus
 1 (bases 1 to 899)
 Boardman, P.E., Bonfield, J.K., Brown, W.R.A., Carder, C., Chalk, S.E.,
 Croning, M.D.R., Davies, R.M., Francis, R.M., Grahame, D.V.,
 Hubbard, S.J., Humphrey, S.J., Hunt, P.J., Maddison, M., McLaren, S.R.,
 Niblett, D., Overton, I.M., Rogers, J., Scott, C.E., Taylor, R.G.,
 Tickle, C. and Wilson, S.A.
 Direct Submission
 Submitted (02-FEB-2004) Sanger Institute, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: chickens@hinxton.ac.uk
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST Gallus gallus cDNA
 sequencing project.
 This sequence is from the
 BBSRC/Dundee/Nottingham/Sanger/Sheffield/UMIST cDNA collection,
 from a library constructed by Elizabeth Bosch. cDNA was prepared
 from RNA extracted from small intestine, normalised, and poly A-tailed.
 EcoRI-NotI cut cDNA was then ligated into the vector.
 Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI Host:
 Escherichia coli DH10B.
 Location/Qualifiers
 1..899
 /organism="Gallus gallus"

FEATURES
 source

Db 529 MEMAGTMMWRHNNNNNTDTRYYMMKRMABRTTYYDSMCKNSMWRGNMRRAMKMMMAA 470

QY 1481 GTGACCCCTCTCTGCAAGATCATGCAAGCTTGAAGCTTGAAGCAGCATCATCA 1540

Db 469 NNDAGANDHTTMMGNNTMMRRAMRMNNAMKRRAYCCNNNNRRACVHKKHMMRWTK 410

QY 1541 TGTAAATTCCTTCTGTAAGCCTGTGATTAATCTCTTTGCAAGCTCTGCTATTA 1600

Db 409 YMKKACNNNNBRAMMYRVAMMYSRDTNTIDMMMTSLMBHMTWYVDYTPMRAMNNNN 350

QY 1601 CCAAGAGTCTGCTACAGACAGAAAGATGCGCTTTAAGCTTCTGAAATAGAGAAA 1660

Db 349 NMMRBCKTTSMMWMDHNMHTCTYGNNTGSAIBMAMMAASNBVTYVMCMRTM 290

QY 1661 ATGAGAACTATGAAAT--TCATGAAGCAATGCGTAAGCCTTACTTTTGAATGCTT 1719

Db 289 GKTMTNNNNNNKMYRTKTVACNNRYDDTAVMTBKRYKCYAVBYBYBMWKHMM 230

QY 1720 CTAAAGCAGATACACCTACGCGGGAATACAAACTTAAGCTGTTCATGAACCATCT 1779

Db 229 BMMRBABHSMMWMMVKCNKMYVSMHYHAMRYBKABAVGNNNNKMDMAHHMWCATNN 170

QY 1780 GGTAACTGTCTCTTGTGTCTGCTATTTTGGCTTGACATTTGCTGCACTTATTT 1839

Db 169 NMMWMMYAVMMHMKKGAAMTNNKTABDDBAHVKTYYMYRYYCAMCMWNAKAYR 110

QY 1840 GAAAGACTCTATAGAGGGGAATACAGAGAAAGAAACATTCGATTTTATTTGC 1894

Db 109 TANKHMYTDRYVSANTGVMMRMCMWMTSMNNRYMRYYMRKRYTAMMYSM 55

RESULT 6

LOCUS RATEFABP 3013 bp DNA linear ROD 27-APR-1993

DEFINITION Rat intestinal fatty acid binding protein gene, exons 1 and 2, and a B2 repetitive element.

ACCESSION M18080 J03465

VERSION M18080.1 GI:204071

KEYWORDS B2 repetitive sequence; fatty acid binding protein.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 3013)
Sweetser, D.
Unpublished (1987)

JOURNAL 2 (bases 917 to 1217)
Sweetser, D.A., Birkmeier, E.H., Klisak, I.J., Zollman, S.,
Sparkes, R.S., Mohandas, T., Lusis, A.J. and Gordon, J.I.
The human and rodent intestinal fatty acid binding protein genes. A
comparative analysis of their structure, expression, and linkage
relationships

TITLE J. Biol. Chem. 262 (33), 16060-16071 (1987)

JOURNAL MEDLINE 88058967

PUBMED 2824476

COMMENT Original source text: Rat (Sprague Dawley) intestinal epithelial
lining DNA, clone lambda-RIFABP.
Draft entry and computer-readable sequence for [2] kindly provided
by D. Sweetser, 19-JAN-1988.

FEATURES

source location/Qualifiers

1. 3013

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/repeats="406

/note="B2 repeat"

prim_transcript 1155..>3013

/note="FABPI mRNA (alt.) and introns"

prim_transcript 1179..>3013

/note="FABPI mRNA (alt.) and introns"

join(1215..1281,2541..2713)

/note="fatty acid binding protein"

/codon_start=1

/protein_id="AAA41133.1"

/db_xref="GI:204072"

/translation="MADPTGKVRNENYKEMKMGITNVYRKUGAHNDLKLITIOE
GNKRTVKSNNFRNIDVFEIGVDPAISLAGTELT"

<1215..1281

/note="fatty acid binding protein"

/number=1

1282..2540

/note="FABPI intron A"

2541..2713

/note="fatty acid binding protein"

/number=2

2714..>3013

/note="FABPI intron B"

ORIGIN 1 bp upstream of EcoRI site.

Query Match 2.6%; Score 62; DB 10; Length 3013;

Best Local Similarity 59.3%; Pred. No. 0.0085;

Matches 127; Conservative 0; Mismatches 80; Indels 7; Gaps 1;

QY 1515 TTGAGCTTTAGCGACCATATCATGTAATTGCTTCTCGATTAAGCCTTCTATAA 1574

Db 1096 TTGAACCTTGAACCTTCACATCATGTAATGATTTGGAAGATAGAAATAGATAA 1155

QY 1575 TTCTCTTGCAAGCTGTGCTACTTACGAGAGTC-----TGCTACAGACAGAAAGA 1627

Db 1156 TTCTCTTAGTGAAGACGACCAATCTGCTTCTTAGAGGACACACAGCTGACATCA 1215

QY 1628 TGGCATTTAAACGCTACTTGGAAAATAGAGAAAATAGAGAAATTCATGAGAG 1687

Db 1216 TGGCAATTTGATGGCACTTGGAAAATAGAGACCGAAATAGAGAAATTCATGAGAG 1275

QY 1688 CAATGGTAAGCCTTACTTTTGAATGCTTCT 1721

Db 1276 AATGGTAAGGCGCTGCTTCTGCTATTTGCT 1309

RESULT 7

LOCUS AC130496

DEFINITION Rattus norvegicus clone CH230-4G2. *** SEQUENCING IN PROGRESS ***

ACCESSION AC130496

VERSION AC130496.4 GI:24818359

KEYWORDS HTGS, HTGS_PHASE1, HTGS_DRAFT, HTGS_ENRICHED.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 230897)

AUTHORS Wuzny, D., Marie, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceaser, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Din, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C.M., Gabiel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, M.,
Gunnarone, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huily, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

CDS
complement(join(10868..10972,11088..11177,11367..11617,
11727..11807,11991..12324))
/gene="PF00335c"
/note="Similar to Plasmodium falciparum hypothetical 33.9
kDa protein mal4p2.04 SMALL:090018 (EMBL:AL035475) (286
aa) fasta scores: E(): 4.9e-117, 98.25% id in 286 aa"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD49160.1"
/db_xref="GI:23498189"
/db_xref="UniProt/TREMBL:O811W3"
/translation="MFNKLVSRTSYFYDTRVNLNKLKLYKQMDYLKRMINY
FNNRKYNTDVIDEAKCWEKFKKELIQANGYPAIIMKXGEERCIHQKIDEPAF
VEEDHLSLFSARFVHIGKVCVSVAGSDVHAIKGIYELFKARVVENQITIN
PESDILGICSPAYINGYVQALMSNICKGVNSNIRGPFQIDVSKLTVPKPNISIKL
KOLLVLPDQGNVIFSEYNLDDETVWVTEPKIMETPLPDDYDNPVNVKGRKIQ
LYKDYWPKO"
CDS
complement(join(12996..13021,13108..13198,13312..13425,
13569..18095,18336..18767,18885..18910,19003..19037,
19326..19336,19432..19509))
/gene="PF00335c"
/note="Alternative start site possible; gene model
slightly modified from submitted version (SMALL:090017
EMBL:AL035475) Plasmodium falciparum hypothetical 205.8
kDa protein sequence
Similar to Plasmodium falciparum hypothetical 205.8 kDa
protein mal4p2.05 SMALL:090017 (EMBL:AL035475) (1734 aa)
fasta scores: E(): 0, 99.25% id in 1742 aa"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD49161.1"
/db_xref="GI:23498190"
/db_xref="UniProt/TREMBL:O811W2"
/translation="NNPFLITLLFPFIMFKLAKGVNIRGMCIDCSPLFENRIS
EDRTVCKSKSEKIDNLDLYNDICTETKLEEDDESCKGFKCQILGTTTHKKN
KVALERKFEEDYLSQFVYKYNNDKNYINNDSKCGVNNISFDNDENPYKND
KNKFINPLEYKKGVEYITPGENVHLITQNGNLIEBKDKLKWSRNKKKY
KYCMNDVSYNS INKKRNINLVNKKLRKKTGYDKNEKPFKRSYNMSKREK
ALSKAKKREKMKKKKKREKKEKESLNVEMENLNCVTKSKENIYNIVENLKL
NEKLEKIQKYNHNDTSYQKLRKREKNDSTIKIEIRKKMDRIYNDKDKDKQI
NVSSLVLEKQSNILKDSKEFLCSNMBIPVNNREKCEKIEKQYIAGSTNTDLYK
DOISNEDSLKKNNE INNESGVNI IKKIKTRSYSDNIDKSDYKKNQVNDYK
VYVNDYSQSCNEKERSYNNYNDYSQSCNEKERSYNNYNDYNDYNNIVE
YNNVSHDYSYNEKGNIIIPYDNIITEENS INKENDDLTOMEOLOSKEIKES
NEDDKDNTKFTNLSHSIKERIKIODINDMEINLPEENKKNVSHVVEKK
NPSSKXKVDNRINKKKNQNIINIDIPNIPYMLILIOHEKTKNKKKREKRY
NIVBELGLPANCVLKALIEKRLFSRIEKLIDVCNMYKONICFRKNKMKRY
NNIVSDDEKRYGPIVSSRSYIPFNNVDKNNIHI KTYNNNYLLNFDGKNDLYN
NNGEERKNVPLIKPLANTINVANEVNTKNNYINKCIYKKKQNDKINLSGIYD
KALNKGIKSYLKKNNINVSKKSIKQDLIKYNNMYSHLKNKCIKTSMLSLC
NMKQSDNVCKNNIYKVNHNKACNKKKESASIKYNNIHNKHNIGINRSTHS
DYKVDISITNGLSIDKDNKGVIIISKVLFMDNTHKEGMYTKRSIMPRNGLKIKIYN
KKRBEKSTFCETKNNDCQVATNEMKAGNDSLQNETIIEKFPLOJENRERIE
EEKQICIVDDKNNINNVDEKRSDDHPSYRVLKMGSGNNEGYSNTQIKLENEKN
KYNNEKKGENDENKKEKENDENKKNYNEKNDENKENDENNNNNSYFVNS
DTEFLCTNSLIFINNKKNSILIPSENEKGIIGQKEBQNIIPKINNKKDLCKIN
ESDYSDKQSVLNSIEKIKYKCSNSKIRGIEKKKINEDYVDLKNINSHYTLFF
LTKKYLKSSRLIINBDCNNINNVKSKKRGAKKKLRKINVINIPNDSIIEENS
EYNPVKKKNNCWFETTKRSKSIISSEIFAIVKKNKKRATNLRSEQFISIGLVKK
GENKRIIEKDEBEYIEKIKNNKNEFPKNLTQLLPFKSAENINTSGSPTEKIRHVX

Query Match 2.2% Score 51.8; DB 3; Length 234081;
Best Local Similarity 45.6% Pred. No. 0.65;
Matches 222; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

QY 1834 TAATTTGAAAAGCTATAGAGGGAATACAAGAAAGAAAACATTTCATTATTATTTG 1893
DB 156675 TAT 156734
QY 1894 CATTGCATATCTTATGCACTTATGCTAATTCAGTAGAGGCATTCCAGCAGAAATTAA 1953
DB 156735 TATACAT 156794

QY 1954 ATAGATATATATATGTAAGGAATATATATTTTGTATAGACTGTTGAAAAATTACACAAGAGG 2013
DB 156795 ATACGAT 156854
QY 2014 GAAATTCGTGCTCCAGCTTTTGCAGAACACACATGATTTAGTCATTTTAAACATGCTNG 2073
DB 156855 ACACCTTTTTTTTTT--TTTAAAGTATATAAAAAGTGTTTATTTATATATATGTTATA 156911
QY 2074 TCGTACTTAAAGCTTGACAAACGCGCTGATATNGATGTAACATATACATCTCTACT 2133
DB 156912 TATATTAATTAATGAATGATATATATATATATATATATATATATATATATATATAT 156971
QY 2134 TCGATATAGTATTTGATATTAACAGCTGAACACTGCTCAGTGAAGTGAGAGAGTAAAG 2193
DB 156972 TTGTTATGTTTATATGAGAAATTTAAATATTTCTTATATATATATAGAAATTTTAC 157031
QY 2194 ACTGAGTCAGAAATTCGGGCTAAAGCTCCGCAACTACAGAAAAAGTCACATATAAAT 2253
DB 157032 GATATATATTTACATGATGTAATATATATATATATATATATATATATATATATAT 157091
QY 2254 GGAACATGATGTTGATTTGTTTCTGCTGCTGATGTTGATTTGATTTATTTATTTT 2313
DB 157092 TTAAT 157151
QY 2314 TTTTATTT 2320
DB 157152 TTTTATTT 157158

RESULT 11
RATFABPX 564 bp mRNA linear ROD 27-APR-1993
LOCUS Rat intestinal fatty acid binding protein (FABP) mRNA, complete
DEFINITION cds
ACCESSION M35992
VERSION M35992.1 GI:204087
KEYWORDS fatty acid binding protein.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 564)
Gordon, J.I. and Lowe, J.B.
Analyzing the structures, functions and evolution of two abundant
gastrointestinal fatty acid binding proteins with recombinant DNA
and computational techniques
Chem. Phys. Lipids 38 (1-2), 137-158 (1985)
86053743
3840724
Original source text: Rat adult small intestinal epithelium, cDNA
to mRNA.

FEATURES

source

1..564
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
13..411
/note="fatty acid binding protein (FABP)"
/codon_start=1
/protein_id="AAA41141.1"
/translation="MAFDGTVVDNENYEKEMKGINVVRKLGADHNLKLTITQE
GNKPTVKESNFRNIDVYVELGVDPAYSLADGTETLGTWTMGNTLVGKFRVNDQKE
LIAVRBISGNEILITTYTEGVAKRIFKKE"

ORIGIN

Query Match 2.1% Score 50; DB 10; Length 564;
Best Local Similarity 75.6% Pred. No. 3.9;
Matches 62; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 1615 ACAGACGAAAGATGCGATTTAACGTTACTTGAAAAATGAGAAAAATGAGAACTATGAA 1674

Db 1 ACAGTCGATCATCATGCGATTGATGCGACTTGGAAGAACCGGAATGAGAACTATGAA 60
 QY 1675 AAATTCATGAGCAATGGCTA 1696
 Db 61 AAGTTCATGAGAAATGGCA 82

RESULT 12
 XELIFABP 619 bp mRNA linear VRT 16-MAR-1994
 LOCUS Xenopus laevis intestinal fatty acid binding protein mRNA, complete cds.
 DEFINITION
 ACCESSION L19946.1 GI:311053
 VERSION 119946.1
 KEYWORDS fatty acid; fatty acid binding protein; hormone-dependent regulation; metamorphosis; thyroid hormone; thyroid hormone dependent.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; 1 (bases 1 to 619)
 REFERENCE Shi, Y.B. and Hayes, W.P. Thyroid hormone-dependent regulation of the intestinal fatty acid-binding protein gene during amphibian metamorphosis Dev. Biol. 161 (1), 48-58 (1994)
 JOURNAL MEDLINE 94123868
 PUBMED 8293885
 REFERENCE 2 (bases 1 to 619)
 AUTHORS Shi, Y.B.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUN-1993) Yun-Bo Shi, Laboratory of Molecular Embryology, Building 6, Rm. B1A02, NICHD/NIH, Bethesda, MD 20892, USA

COMMENT Original source text: Xenopus laevis cDNA to mRNA.
 FEATURES
 source location/Qualifiers
 1..619
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /cell_type="epithelial cell"
 /tissue_type="intestine"
 /dev_stage="tadpole"
 /tissue_lib="Lambda Zap"
 41..439
 /codon_start=1
 /product="fatty acid binding protein"
 /protein_id="AAC38012.1"
 /db_xref="GI:311054"
 /translation="MAPDGTWKVDRESEYKEMVWGCVNIYKRIKAHNLKVTIQQD GNNFTVSESTFNNIRIKFTLQAPFYSIADGELGAWFLQDNOLLGTFTRDKNGKVLQTRROIIGDELQVTEYEETSEKRIKFG"

ORIGIN

Query Match 2.1%; Score 49.8; DB 5; Length 619;
 Best Local Similarity 68.3%; Pred. No. 4.2;
 Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1595 TACTTACGAGAGTCTGCTACAGACAGAAAGATGATTACCGGACTTGGAAATG 1654
 Db 9 TGGCTAGCCAAAGCCTCTCTCCACCAATCATGCGCTTGTATGGAAGGTTG 68

QY 1655 AGAAAAATGAGAACTATGAAAAATTCATGGAAGCAATGGCT 1695
 Db 69 ACAGAAAGTGAAGAACTATGAGAAATTCATGGAAGTTAGGT 109

RESULT 13
 LOCUS AX771590 5204 bp DNA linear PAT 02-JUL-2003
 DEFINITION Sequence 303 from Patent WO03004646.
 ACCESSION AX771590

VERSION AX771590.1 GI:32438388
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 O'Mahony, D.J., Byrne, D., Brayden, D., Lambkin, I., and Higgins, L. Genetic analysis of Peyer's patches and M cells and methods and compositions targeting Peyer's patches and M cell receptors Patent: WO 03004646-A 303 16-JAN-2003;
 ELAN CORPORATION, PLC (IE)

FEATURES
 source location/Qualifiers
 1..5204
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 2.1%; Score 49.8; DB 6; Length 5204;
 Best Local Similarity 59.2%; Pred. No. 3.1;
 Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;

QY 1516 TGAGCTTTAGCCGACCATCATCATGTAATGCTTTCCTGATAGCCGTTCTATTAAT 1575
 Db 971 TGAACCTTAAGCTTCCACATCATCAGATGAGTTGGTTCAAGTAAAGAAATTAATTAAT 1030

QY 1576 TCTC-----TTTGCAAGCTCTGCTACTTACCAAGAGTGCCTACAGAGAAAGA 1627
 Db 1031 TCTCGCCCAAGAGACACACCTGATCTCTAGCTGCTTAGAGGCTGACT-CACTGAAATCA 1089

QY 1628 TGGCATTTAACGGTACTTGGAAATAGAGAAATAGAACTATGAAAAATTCATGGAAG 1687
 Db 1090 TGGCGTTTACACGACTTGGAAAGTACCGGAGTAAAACTATGACAGGTTATGAGAAA 1149

QY 1688 CAATGGTAAGCCTTACTTTTGAATGCCT 1718
 Db 1150 AAATGGTAAGACCTTATTTCTTTGTGGCT 1180

RESULT 14
 HUMFABP 5204 bp DNA linear PRI 08-NOV-1994
 LOCUS Human, intestinal fatty acid binding protein gene, complete cds, and an Alu repetitive element.
 DEFINITION
 ACCESSION M18079.1 GI:182351
 VERSION M18079.1
 KEYWORDS Alu repeat; fatty acid binding protein.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 5204)
 Sweetser, D.A., Birkenmeier, E.H., Klisek, I.J., Zollman, S., Sparkes, R.S., Mohandas, T., Luisis, A.J., and Gordon, J.I. The human and rodent intestinal fatty acid binding protein genes. A comparative analysis of their structure, expression, and linkage relationships J. Biol. Chem. 262 (33), 16060-16071 (1987)

JOURNAL Original source text: Human DNA (library of T.Maniasis), clone lambda-HIFABP.
 MEDLINE Draft entry and computer-readable sequence for [1] kindly provided by D.Sweetser, 19-JAN-1988.
 PUBMED 2824476

COMMENT

FEATURES
 source location/Qualifiers
 1..5204
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /map="4q28-q31"
 prim_transcript 1028..54393
 /note="FABPI mRNA (alc.) and introns"

	prim_transcript	1053..>4393	/note="FABPI mRNA (alc.) and introns"
CDS		join(1089..1155,2350..2522,3546..3653,4098..4148)	
		/note="fatty acid binding protein"	
		/codon_start=1	
		/protein_id="AA52417.1"	
gene		</translation>"MAFDSTWIKVDRSSNDYKEMKQGVNIYVKRKLAAHNLKLTIOE GNKTIVRESSAFRNIEVFEIAGTFENLMDGTELRGTWSLBNKILIGKEPRDNGBE LMTRREIIDEELVTGYVEGEVARRIPEKO"	
		1089..1155	/gene="FABP2"
exon		<1089..1155	/gene="FABP2"
		/note="fatty acid binding protein; G00-119-127"	
		/number=1	
intron		1156..2349	/note="FABPI intron A"
		2350..2522	/number=2
exon		2523..3545	/note="FABPI intron B"
		3546..3653	/number=3
exon		3654..4097	/note="FABPI intron C"
intron		4098..>4148	/note="fatty acid binding protein"
exon		/number=4	
repeat_region		4466..4742	/note="Alu repeat"
repeat_region		4466..4472	/note="5' direct repeat"
repeat_region		4736..4742	/note="3' direct repeat"
ORIGIN	156 bp upstream of HindIII site; chromosome 4q28-q31.		
Query Match	2.1%; Score 49.8; DB 9; Length 5204;		
Best Local Similarity	59.2%; Pred. No. 3.1;		
Matches 125; Conservative	0; Mismatches 77; Indels 9; Gaps 2;		
OY	1516 TGAGCTTTAGCCAGGCACATCATCGTAAATTGGTTTCGTATPAGCCTTGCAATAAT	1575	
Db	971 TGAAGTTTAAGCTTCACATCACAGTAGAAGTTGGTCAAGATAGAAAATATATATAAT	1030	
OY	1576 TCTC-----TTTGCAAAGTCGTGCTACTTACCAGAAGCTGCTACAGACAGAAAGA	1627	
Db	1031 TCTGCCCCAGGACAGACCTGATCTTCTAGCTGCTTAGAGGCTGACT-CAACTGAAATCA	1089	
OY	1628 TGGCATTTAACGGTACTTGGAAAAATAGAGAAAAATGAGAACTATGAAAAATTCATGAG	1687	
Db	1030 TGGCTTTGACAGCACTTGGAAGTAGACCGGAGTGAATAACTATGACAAGTTTCATGAAA	1149	
OY	1688 CAATGGTAPAGCCTTACTTTTTTGAAAGCTT	1718	
Db	1150 AAATGGTAAAGACTTATTTCTTGTGGCT	1180	
RESULT 15			
LOCUS	AC008077.c		
DEFINITION	Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 20		
ACCESSION	AC008077		
VERSION	AC008077.1 GI:5540103		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 200000) Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M. Direct Submission		

Journal Reference Authors Title Journal Comment	Unpublished 2 (bases 1 to 200000)	Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M. Direct Submission Submitted (20-JUL-1999) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
	* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
	* 1 1146: contig of 1146 bp in length	
	* 1147 10118: gap of unknown length	
	* 10119 11308: contig of 1190 bp in length	
	* 11309 20280: gap of unknown length	
	* 20281 21488: contig of 1208 bp in length	
	* 21489 30460: gap of unknown length	
	* 30461 31634: contig of 1174 bp in length	
	* 31635 40606: gap of unknown length	
	* 40607 41979: contig of 1373 bp in length	
	* 41980 50951: gap of unknown length	
	* 50952 52049: contig of 1098 bp in length	
	* 52050 61021: gap of unknown length	
	* 61022 62105: contig of 1084 bp in length	
	* 62106 71077: gap of unknown length	
	* 71078 72391: contig of 1314 bp in length	
	* 72392 81363: gap of unknown length	
	* 81364 82566: contig of 1203 bp in length	
	* 82567 91538: gap of unknown length	
	* 91539 92900: contig of 1362 bp in length	
	* 92901 101871: gap of unknown length	
	* 101872 103074: contig of 1203 bp in length	
	* 103075 112045: gap of unknown length	
	* 112046 113302: contig of 1257 bp in length	
	* 113303 122273: gap of unknown length	
	* 122274 123778: contig of 1505 bp in length	
	* 123779 132749: gap of unknown length	
	* 132750 134818: contig of 2069 bp in length	
	* 134819 143789: gap of unknown length	
	* 143790 145506: contig of 1717 bp in length	
	* 145478 154477: gap of unknown length	
	* 154478 156731: contig of 2254 bp in length	
	* 156732 165702: gap of unknown length	
	* 165703 167355: contig of 1653 bp in length	
	* 167356 176326: gap of unknown length	
	* 176327 178351: contig of 2025 bp in length	
	* 178352 187322: gap of unknown length	
	* 187323 188990: contig of 1666 bp in length	
	* 188991 197961: gap of unknown length	
	* 197962 200000: contig of 2039 bp in length.	
FEATURES	Location/Qualifiers	
source	1. 200000	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="4"	
ORIGIN		
Query Match	2.1% Score 49.8; DB 2; Length 200000;	
Best Local Similarity	59.2% Pred. No. 1.8; 77; Indels 9; Gaps 2;	
Matches 125; Conservative	0; Mismatches	
QY	1516 TGAAGCTTACGCGCCACATCATCATGTAATGCTTTCTCGATAGACCTGTTCAAT 1575	
DB	10910 TGAAGCTTACGCGCCACATCATCATGTAATGCTTTCTCGATAGACCTGTTCAAT 10851	
QY	1576 TCTC-----TTGCAAGCTCTGCTACTTACGAAAGTCTGCTTACGACAGAGAA 1627	
DB	10850 TCTGCGCCAGACGACAGCTGATCTCTGTAGCTGCTTACGAGGCTGACT-CAACTGAAATCA 10792	

QY 1628 TGGCATTTAACGGTACTTGGAAAAATAGAAAAATGAACTATGAAAAATTCATGGAG 1687
Db 10791 TGGCGTTTGACAGCACTTGGAGGTAGACCGAGTGAAAACTATGACAGTTCTATGSA 10732
QY 1688 CAATGGGTAAAGCCTTACTTTTGAATGCGT 1718
Db 10731 AAATGGGTAAAGACTTATTCTTGTGGCT 10701

Search completed: November 21, 2005, 00:03:22
Job time : 9980 secs

Time Age Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 15:37:53 ; Search time 1191 Seconds
(without alignments)
11834.512 Million cell updates/sec

Title: US-10-099-663-1

Sequence: 1 agcttcctcgccgcagaaagg.....atctgaagctcactatcag 2381

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001bs:*
5: geneseqn2001bs:*
6: geneseqn2002bs:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2381	100.0	2381	11	ADL90127	ADL90127 Chicken i
2	336	14.1	236	11	ADL90128	ADL90128 Chicken i
3	50.6	2.1	4590	5	AAH24065	Aah24065 Yeast ADP
4	49.8	2.1	5204	6	ABK83946	Abk83946 Human CDN
5	49.8	2.1	5204	12	ADJ74588	Adj74588 Human int
6	49.8	2.1	5204	12	ADO78215	Ado78215 Human fat
7	49.2	2.1	2000	8	ADA71938	Ada71938 Rice gene
8	47.2	2.0	337	8	ABK40871	Abk40871 Bovine ES
9	47	2.0	2636	4	ABL15880	Ab115880 Drosophila
10	46.8	2.0	662	10	ADD29587	Add29587 Mouse tum
11	46.6	2.0	610	4	AA119953	Aa119953 Human bre
12	46.6	2.0	9742	6	ABL70479	Ab170479 Chemical
13	46.2	1.9	6222	6	ABL32692	Ab132692 Human imm
14	45	1.9	10652	6	ABN80056	Abn80056 Human che
15	44.4	1.9	2000	8	ADA71938	Ada71938 Rice gene
16	44.2	1.9	515	3	AAAI6469	Aaa16469 Human col
17	44.2	1.9	2252	13	ADQ39082	Adq39082 Human SNP
18	43.6	1.8	14023	6	ABL34104	Ab134104 Human imm
19	43.2	1.8	7143	6	ABL32983	Ab132983 Human imm
20	42.8	1.8	51289	13	ABD33486	Abd33486 Human can

C 21	42.6	1.8	513	12	ACH79811	Ach79811 Human gen
C 22	42.6	1.8	611	3	AAA16371	Aaa16371 Human col
C 23	42.6	1.8	2844	10	ABT31903	Abt31903 Human bre
C 24	42.6	1.8	5501	6	ABJ32055	Abj32055 Human imm
C 25	42.4	1.8	230	2	AAAX1681	Aax11681 Human bia
C 26	42.4	1.8	230	2	AAAX1293	Aax1293 Human dia
C 27	42.4	1.8	452	3	AAAT7858	Aaa7858 cDNA enco
C 28	42.4	1.8	452	4	AAI78596	Aai78596 Colon tum
C 29	42.4	1.8	452	8	ABZ32782	Abz32782 Human col
C 30	42.4	1.8	461	4	AAI10372	Aai10372 Human bre
C 31	42.4	1.8	682	11	ACN80571	Acn80571 Breast ca
C 32	42.4	1.8	5908	6	AA545387	Aaa545387 Chemical
C 33	42.4	1.8	5908	6	ABK28322	Abk28322 DNA trans
C 34	42.2	1.8	453	6	ABE67809	Ab167809 Oesophagu
C 35	42.2	1.8	453	6	ABE62188	Ab162188 Colon ade
C 36	42.2	1.8	6621	6	ABJ32912	Abj32912 Human imm
C 37	42	1.8	9095	6	ABO67061	Abg67061 Human ang
C 38	41.8	1.8	10048	6	ABL70313	Ab170313 Chemical
C 39	41.8	1.8	10048	6	AA561251	Aaa61251 Human gen
C 40	41.8	1.8	137870	10	ADG89426	Adg89426 Human mat
C 41	41.6	1.7	900	4	AAH34536	Aah34536 Human col
C 42	41.6	1.7	3639	6	ABE62356	Ab162356 Colon ade
C 43	41.6	1.7	3639	12	ADJ74884	Adj74884 Marker ge
C 44	41.6	1.7	3719	9	ADA10896	Ada10896 Human CDN
C 45	41.4	1.7	447	8	ABX50540	Abx50540 Bovine ES

ALIGNMENTS

RESULT 1	ADL90127	standard; DNA; 2381 BP.
ID	ADL90127	
AC	ADL90127;	
XX		
DT	20-MAY-2004	(first entry)
XX		
DE	Chicken intestinal fatty acid binding protein, IFABP, gene, 5' region.	
XX		
KW	Chicken; ds; intestinal fatty acid binding protein; IFABP;	
XX	gut specific promoter; transgenic.	
OS	Gallus gallus.	
XX		
PN	US2003177516-A1.	
XX		
PD	18-SEP-2003.	
XX		
PF	14-MAR-2002; 2002US-00099663.	
XX		
PR	14-MAR-2002; 2002US-00099663.	
XX		
PA	(HORS/) HORSEMAN N D.	
XX		
PI	(PRAT/) PRATT S L.	
XX		
PI	Horseman ND, Pratt SL;	
XX		
DR	WPI; 2003-898653/82.	
XX		
PT	New nucleic acid molecule comprising an isolated avian gut-specific gene	
XX	expression control region, useful for regulating heterologous nucleic	
PT	acids in transgenic avians, and for generating transgenic birds.	
XX		
PS	Claim 1; SEQ ID NO 1; 28bp; English.	
XX		
CC	The invention relates to an isolated nucleic acid comprising an isolated	
CC	avian gut-specific gene expression control region appearing as	
CC	ADL90127(Chicken intestinal fatty acid binding protein, IFABP, gene, 5'	
CC	region or ADL90128 (Chicken IFABP promoter) or its degenerate variant.	
CC	Also included are a recombinant DNA molecule comprising an isolate avian	
CC	gut-specific gene expression control region operably linked to a nucleic	
CC	acid insert encoding a polypeptide, an expression vector that integrates	

CC into a host cell (and comprising the isolated avian gne-specific gene
CC expression control region), expressing a heterologous polypeptide in a
CC host cell (by transfecting a eukaryotic cell with the recombinant DNA
CC molecule, and culturing the transfected cell in a medium suitable for
CC expression of a heterologous polypeptide under the control of an avian
CC intestinal fatty acid binding protein (IFABP) or cp15 gene expression
CC control region encoded by the recombinant DNA molecule), a eukaryotic
CC cell transformed with the expression vector (or its progeny, which
CC expresses a heterologous polypeptide) and a transgenic avian having a
CC heterologous polynucleotide sequence comprising the nucleic acid insert.
CC The nucleic acids are useful for regulating heterologous nucleic acids in
CC transgenic avians, as probes in nucleic acid hybridization assays for
CC detecting the IFABP gene expression control region, and for generating
CC transgenic birds. The present sequence is the Chicken intestinal fatty
CC acid binding protein, IFABP, gene, 5' region.

XX Sequence 2381 BP; 775 A; 428 C; 469 G; 709 T; 0 U; 0 Other;

Query Match 100.0%; Score 2381; DB 11; Length 2381;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCTTCCTGGCGAAGAAAGGCTGTGGGTTCTTGTTCCTCAACAGCTTAAAGCAAAATC 60
DB 1 AGCTTCCTGGCGAAGAAAGGCTGTGGGTTCTTGTTCCTCAACAGCTTAAAGCAAAATC 60
QY 61 CCCAAGTTCAAAACGTCGCTGTGAAGAGATGCGTCACTTCAATGAAGTAATG 120
DB 61 CCCAAGTTCAAAACGTCGCTGTGAAGAGATGCGTCACTTCAATGAAGTAATG 120
QY 121 AAATATCATTAACAGAGCTCTGTGGCAGATCAGAGATTAACCTGTGCTGGGCAAAAT 180
DB 121 AAATATCATTAACAGAGCTCTGTGGCAGATCAGAGATTAACCTGTGCTGGGCAAAAT 180
QY 181 CTTAAAGTGTATAGTAAAGAGAGGCTGTGGCACTAAATGAGTAAATGTACACA 240
DB 181 CTTAAAGTGTATAGTAAAGAGAGGCTGTGGCACTAAATGAGTAAATGTACACA 240
QY 241 TTTGATCTTCTAGGAGACAAAGGCTGTGGAAACAAATTAATCTGTGTGCACAGTCA 300
DB 241 TTTGATCTTCTAGGAGACAAAGGCTGTGGAAACAAATTAATCTGTGTGCACAGTCA 300
QY 301 AGCAGCTGTTTTGGGTGCACTACAGCACTTTGTTGGCAATTAACATTTAAGTTG 360
DB 301 AGCAGCTGTTTTGGGTGCACTACAGCACTTTGTTGGCAATTAACATTTAAGTTG 360
QY 361 TTTTCTTTTCTCTTCTTAACTCTGTACAGCTTAAAGGTAAAGTAGTATTGAGT 420
DB 361 TTTTCTTTTCTCTTCTTAACTCTGTACAGCTTAAAGGTAAAGTAGTATTGAGT 420
QY 421 TACTTCCCTCTGATCCTCTTAAGCAGATTAGCATTTTCAAAATGAACCTGAGTGA 480
DB 421 TACTTCCCTCTGATCCTCTTAAGCAGATTAGCATTTTCAAAATGAACCTGAGTGA 480
QY 481 ATGGAAGACACACTATTTTGTATCAACAGAAAGTTCTAAATTTATAGTTACTTCA 540
DB 481 ATGGAAGACACACTATTTTGTATCAACAGAAAGTTCTAAATTTATAGTTACTTCA 540
QY 541 GTAAGACCTTTTCTGAGGCTGTGAAGAAAGAAATTAATATACACAGACTAGTA 600
DB 541 GTAAGACCTTTTCTGAGGCTGTGAAGAAAGAAATTAATATACACAGACTAGTA 600
QY 601 AAATTCATTAGTAGAGCCAACTGTTATCTGTGTGATAGCAATTCATTTCAAGCAT 660
DB 601 AAATTCATTAGTAGAGCCAACTGTTATCTGTGTGATAGCAATTCATTTCAAGCAT 660
QY 661 TCAGAGATTTCATTTTGAAGCTTAATAGACAGAGATTTGTGCGCTCAATAGAACAGA 720
DB 661 TCAGAGATTTCATTTTGAAGCTTAATAGACAGAGATTTGTGCGCTCAATAGAACAGA 720
QY 721 CTAACATTAATCTGAGATTAGTACAAGAGATTTAGCACAGCAATTTGCTCAAGTTTC 780
DB 721 CTAACATTAATCTGAGATTAGTACAAGAGATTTAGCACAGCAATTTGCTCAAGTTTC 780

QY 781 AAGTAGCACTAATCTGTGGGGAAGAGAGGCTGAGCCAGTGTGCTCATTTTCGACT 840
DB 781 AAGTAGCACTAATCTGTGGGGAAGAGAGGCTGAGCCAGTGTGCTCATTTTCGACT 840
QY 841 ATCTTCAACATTTTAAACCTGGGATCTATGGAATCAACACGTTGGGTAAATTCAC 900
DB 841 ATCTTCAACATTTTAAACCTGGGATCTATGGAATCAACACGTTGGGTAAATTCAC 900
QY 901 TAGCGACATCACTACTGTGTGGAATGACAGAAACAGAGATTCAGTAATGGGCTAT 960
DB 901 TAGCGACATCACTACTGTGTGGAATGACAGAAACAGAGATTCAGTAATGGGCTAT 960
QY 961 AATATAGGAATAGCTAGGAAGGTGCTGAATTTAGACTACCTATTTAAAGTAGAGACA 1020
DB 961 AATATAGGAATAGCTAGGAAGGTGCTGAATTTAGACTACCTATTTAAAGTAGAGACA 1020
QY 1021 CGAATGAGAAATATCATGCAATTTCTGTAGCTCAGCACTAGACTCGAAGTTTCTGAAA 1080
DB 1021 CGAATGAGAAATATCATGCAATTTCTGTAGCTCAGCACTAGACTCGAAGTTTCTGAAA 1080
QY 1081 CTGAACCGAGTTTCCCAAACCTACCTGTGAGATGTTCAATCTCATCTCAT 1140
DB 1081 CTGAACCGAGTTTCCCAAACCTACCTGTGAGATGTTCAATCTCATCTCAT 1140
QY 1141 TATGTGAGTGAATAGATTTCTACCAAAATTAAGATGAAGAGAGAGATTTGTTT 1200
DB 1141 TATGTGAGTGAATAGATTTCTACCAAAATTAAGATGAAGAGAGAGATTTGTTT 1200
QY 1201 ATCTGTGGGTAAATAGCTTTTCTCAGTGTATTAAGACCTCCACAGATTAAGT 1260
DB 1201 ATCTGTGGGTAAATAGCTTTTCTCAGTGTATTAAGACCTCCACAGATTAAGT 1260
QY 1261 CTATGCAACAAAGAAATGTCAATACATCTCTTATGCTCATTTATTTTATTAAGATA 1320
DB 1261 CTATGCAACAAAGAAATGTCAATACATCTCTTATGCTCATTTATTTTATTAAGATA 1320
QY 1321 GCGGTTTTTTTCTCAACTCAATTAAGTGAAGAAATGAATGGGTTAGTACTGTTA 1380
DB 1321 GCGGTTTTTTTCTCAACTCAATTAAGTGAAGAAATGAATGGGTTAGTACTGTTA 1380
QY 1381 TAAAGAGAGTAAATTAAGATATCTATCATCATTTTGAAGCAATTAAGGAGAGAGATT 1440
DB 1381 TAAAGAGAGTAAATTAAGATATCTATCATCATTTTGAAGCAATTAAGGAGAGAGATT 1440
QY 1441 CAAACAGTGTCTTACAGTGAAGAAACAAATTAAGTGAAGTCCCTCTTGACAA 1500
DB 1441 CAAACAGTGTCTTACAGTGAAGAAACAAATTAAGTGAAGTCCCTCTTGACAA 1500
QY 1501 GATCAATGCAAGTTGAGCTTTAGCGAGCCACATCATGATTAATGCTTCCGAT 1560
DB 1501 GATCAATGCAAGTTGAGCTTTAGCGAGCCACATCATGATTAATGCTTCCGAT 1560
QY 1561 AGCTGTTCTAATAATCTTCTTGAAGAGCTGTCTACTTACAGAGAGTGTGCTACAG 1620
DB 1561 AGCTGTTCTAATAATCTTCTTGAAGAGCTGTCTACTTACAGAGAGTGTGCTACAG 1620
QY 1621 AGAAGATGCACTTTTAAACGTTACTTGAAAAATAGAAAAATGAGAACTATCAAAAATTC 1680
DB 1621 AGAAGATGCACTTTTAAACGTTACTTGAAAAATAGAAAAATGAGAACTATCAAAAATTC 1680
QY 1681 ATGGAAGCAATGGGTAAGCTTACTTTTGAATGCTCTTAAGAGAGATTAACCACTAC 1740
DB 1681 ATGGAAGCAATGGGTAAGCTTACTTTTGAATGCTCTTAAGAGAGATTAACCACTAC 1740
QY 1741 GCGGAAATCAAACTTAAGCTGTTCATGAACTACATCTGGCTAACTCTGCTTGTGT 1800
DB 1741 GCGGAAATCAAACTTAAGCTGTTCATGAACTACATCTGGCTAACTCTGCTTGTGT 1800
QY 1801 CTGCTATTTTGGCTTGCACATTTGCGCTGCACTTATTTTGAAGAAAGCTCATATAGAGGGA 1860
DB 1801 CTGCTATTTTGGCTTGCACATTTGCGCTGCACTTATTTTGAAGAAAGCTCATATAGAGGGA 1860

QY 1861 ATCAAGGAAGAAAATTCATGATTTTATTCAGTGTGCAATATCTTATGCAATTTAGCT 1920
 DB 1861 ATCAAGGAAGAAAATTCATGATTTTATTCAGTGTGCAATATCTTATGCAATTTAGCT 1920
 QY 1921 AATTCAGTAGAGGCAATTCAGAGAGAAATTTAATAGAAATTTATGTAAGCAATATTAAT 1980
 DB 1921 AATTCAGTAGAGGCAATTCAGAGAGAAATTTAATAGAAATTTATGTAAGCAATATTAAT 1980
 QY 1981 TTGATAGAGCTGTTTGAATAATTAACACAGAGAGAAATTCGTCGCTCCAGTTTTCAGA 2040
 DB 1981 TTGATAGAGCTGTTTGAATAATTAACACAGAGAGAAATTCGTCGCTCCAGTTTTCAGA 2040
 QY 2041 ACACAGATGATTTGAGCATTTTAAACATGCTAGTCTTATTTAGCTTTAAGCTTAACTGC 2100
 DB 2041 ACACAGATGATTTGAGCATTTTAAACATGCTAGTCTTATTTAGCTTTAAGCTTAACTGC 2100
 QY 2101 CTGTAAATATGATGATTAACATTAATCTAGTTGATAGTATGTTGTATTAACAGGCTGA 2160
 DB 2101 CTGTAAATATGATGATTAACATTAATCTAGTTGATAGTATGTTGTATTAACAGGCTGA 2160
 QY 2161 ACACTGCTCAGTGAAGAGTGAGAAAGATGAAGTCTGAGTCAAGATTCGGGCTAAGC 2220
 DB 2161 ACACTGCTCAGTGAAGAGTGAGAAAGATGAAGTCTGAGTCAAGATTCGGGCTAAGC 2220
 QY 2221 TCCCTCAGCTAGAGAAAAGTCAATTAATAAGCAATGATTTCTATTTTGT 2280
 DB 2221 TCCCTCAGCTAGAGAAAAGTCAATTAATAAGCAATGATTTCTATTTTGT 2280
 QY 2281 CTCGCTGATGATTTAATGATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2340
 DB 2281 CTCGCTGATGATTTAATGATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2340
 QY 2341 GAAAGTTAGAGCCCAAGATTAATCTGAAGCTCATAATTCAG 2381
 DB 2341 GAAAGTTAGAGCCCAAGATTAATCTGAAGCTCATAATTCAG 2381
 RESULT 2
 ADL90128
 ID ADL90128 standard; DNA; 336 BP.
 AC ADL90128;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Chicken intestinal fatty acid binding protein, IFABP, gene, promoter.
 XX
 KM Chicken; ds; intestinal fatty acid binding protein; IFABP;
 KW gut specific promoter; transgenic; promoter.
 XX
 OS Gallus gallus.
 XX
 PN US2003177516-A1.
 XX
 PD 18-SEP-2003.
 XX
 PF 14-MAR-2002; 2002US-00099663.
 XX
 PR 14-MAR-2002; 2002US-00099663.
 XX
 PA (HORS/) HORSEMAN N D.
 PA (PRAT/) PRATT S L.
 PI Horseman ND, Pratt SL;
 PI
 DR WPI; 2003-898653/82.
 XX
 PT New nucleic acid molecule comprising an isolated avian gut-specific gene
 PT expression control region, useful for regulating heterologous nucleic
 PT acids in transgenic avians, and for generating transgenic birds.
 XX
 PS Claim 1; SEQ ID NO 2; 28pp; English.
 XX

CC The invention relates to an isolated nucleic acid comprising an isolated
 CC avian gut-specific gene expression control region appearing as
 CC ADL90127 (Chicken intestinal fatty acid binding protein, IFABP, gene, 5'
 CC region or ADL90128 (chicken IFABP promoter) or its degenerate variant.
 CC Also included are a recombinant DNA molecule comprising an isolate avian
 CC gut-specific gene expression control region operably linked to a nucleic
 CC acid insert encoding a polypeptide, an expression vector that integrates
 CC into a host cell (and comprising the isolated avian gut-specific gene
 CC expression control region), expressing a heterologous polypeptide in a
 CC host cell (by transfecting a eukaryotic cell with the recombinant DNA
 CC molecule, and culturing the transfected cell in a medium suitable for
 CC expression of a heterologous polypeptide under the control of an avian
 CC intestinal fatty acid binding protein (IFABP) or cp35 gene expression
 CC control region encoded by the recombinant DNA molecule), a eukaryotic
 CC cell transformed with the expression vector (or its progeny, which
 CC expresses a heterologous polypeptide) and a transgenic avian having a
 CC heterologous polynucleotide sequence comprising the nucleic acid insert.
 CC The nucleic acids are useful for regulating heterologous nucleic acids in
 CC transgenic avians, as probes in nucleic acid hybridisation assays for
 CC detecting the IFABP gene expression control region, and for generating
 CC transgenic birds. The present sequence is the Chicken intestinal fatty
 CC acid binding protein, IFABP, gene, promoter.
 XX
 SQ Sequence 336 BP; 116 A; 62 C; 63 G; 95 T; 0 U; 0 Other;
 Query Match 14.1%; Score 336; DB 11; Length 336;
 Best Local Similarity 100.0%; Pred. No. 7.8e-70;
 Matches 336; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1301 ATTAATTTTCAATTTGATGAGCGGTTTCTACTACATCAATTAAGTAAACAGAAAG 1360
 DB 1 ATTAATTTTCAATTTGATGAGCGGTTTCTACTACATCAATTAAGTAAACAGAAAG 60
 QY 1361 AATGGGTAGTACTGTTTATAAGAGATTAATAAGTATCATCATCTTTGAGGCA 1420
 DB 61 AATGGGTAGTACTGTTTATAAGAGATTAATAAGTATCATCATCTTTGAGGCA 120
 QY 1421 TAAGGAGGAGAGATTCAGCAAAACAGTGTCTTACAGTGGAAAACAAGTTAACTAA 1480
 DB 121 TAAGGAGGAGAGATTCAGCAAAACAGTGTCTTACAGTGGAAAACAAGTTAACTAA 180
 QY 1481 GTGACCCCTCTCTTGAACAAGATTCATGCAAGTGGAGCTTTAGCCAGCATCATCA 1540
 DB 181 GTGACCCCTCTCTTGAACAAGATTCATGCAAGTGGAGCTTTAGCCAGCATCATCA 240
 QY 1541 TGTAAATTCCTTCCGATTAAGCTGTTCAATTAATCTCTTGGCAAGCTGTGCTACTTA 1600
 DB 241 TGTAAATTCCTTCCGATTAAGCTGTTCAATTAATCTCTTGGCAAGCTGTGCTACTTA 300
 QY 1601 CCAGAAAGTCTGCTTACAGACAGAAAGATGGCATTTA 1636
 DB 301 CCAGAAAGTCTGCTTACAGACAGAAAGATGGCATTTA 336
 RESULT 3
 AAH24065
 ID AAH24065 standard; DNA; 4590 BP.
 AC AAH24065;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.
 XX
 KM Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
 KM modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
 KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
 KW functional food; transgenic yeast; fat/lean ratio; food use; ds.
 XX
 OS Saccharomyces cerevisiae.
 XX
 FT Key Location/Qualifiers
 FT misc_feature 10

PT	/+tag= a	"Represented as * in the specification"
FT	/+note=	3617
FT	/+tag= b	"Represented as * in the specification"
FT	/+note=	3649
FT	/+tag= c	"Represented as * in the specification"
FT	/+note=	3679
FT	/+tag= d	"Represented as * in the specification"
FT	/+note=	3619
FT	/+tag= e	"Represented as * in the specification"
FT	/+note=	3662
FT	/+tag= f	"Represented as * in the specification"
FT	/+note=	3664
FT	/+tag= g	"Represented as * in the specification"
FT	/+note=	3888
FT	/+tag= h	"Represented as * in the specification"
FT	/+note=	3690
FT	/+tag= i	"Represented as * in the specification"
FT	/+note=	3912
FT	/+tag= j	"Represented as * in the specification"
FT	/+note=	3914
FT	/+tag= k	"Represented as * in the specification"
FT	/+note=	3938
FT	/+tag= l	"Represented as * in the specification"
FT	/+note=	3939
FT	/+tag= m	"Represented as * in the specification"
FT	/+note=	3941
FT	/+tag= o	"Represented as * in the specification"
FT	/+note=	3943
FT	/+tag= p	"Represented as * in the specification"
FT	/+note=	4361
FT	/+tag= q	"Represented as * in the specification"
FT	/+note=	
XX		
PN	WO200133977-A1.	
XX		
XX	17-MAY-2001.	
PD		
XX		
PF	06-NOV-2000; 2000WO-AU001362.	
XX		
PR	05-NOV-1999; 99AU-00003875.	
XX		
XX	(META-) METABOLIC PHARM LTD.	
PA		
XX		
PI	Belyea CI, Ng FM, Vaughan P;	
XX		
DR	WPI, 2001-328976/34.	
XX		
PT	New Organisms containing nucleic acid encoding a growth hormone fragment	
PT	which modulates lipid metabolism are useful to produce dietary aids for	
PT	obesity and in the meat production industry.	
XX		
PS	Disclosure; Page 48-50; 54pp; English.	
XX		
CC	The invention relates to novel transgenic organisms useful in the	
CC	production of functional food and drink products for the treatment or	
CC	prevention of obesity via the regulation of lipid metabolism. The	
CC	organisms comprise a polynucleotide encoding a growth hormone fragment	
CC	capable of stimulating the activity of hormone-sensitive lipase (the key	
CC		

	Query Match	2.1%, Score 50.6, DB 5, Length 4590;	
	Best Local Similarity 11.8%, Pred. No. 0.16;		
	Matches 113; Conservative 416; Mismatches 423; Indels 3; Gaps 3;		
Qy	153 CAGAGATPACTCTGCTGGGACAAATCTTAAAGTATAGTGAACAGAGAGTGT	212	enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key enzyme in lipogenesis). The growth hormone fragment preferably contains at least the disulphide-bonded loop of a mammalian growth hormone (but is not the full-length growth hormone) and is optionally linked to an epitope tag or heterologous fusion protein partner. The transgenic organism may be a microorganism used to produce a fermented product (e.g., yeast), or an edible plant or animal or cell thereof. Food or drink made using methods of the invention are used to modify fat/lean ratio, lipid metabolism or food use in a mammal. In particular, the food or drink products may be used to treat or prevent obesity, particularly in humans, and may also be used to improve the fat/lean ration of livestock raised for meat production. In the exemplification of the invention, the human growth hormone (hgh) fragment analogue A009604 was expressed in yeast, optionally fused to the FLAG epitope (AAB73625). The present sequence is described as a DNA sequence from yeast in the sequence listing, but is not further referred to in the specification
SQ	Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 0 U; 2896 Other;		
Db	3544 CCRNGTCAMSBSTANTAYASHRSDSCBWDTHRRNCTTHAMSANDDRAMNGSDATDTHS	3603	
Qy	213 GCAGCTAAATGAGTATGTTGATACACATTTGATCTCTAGGACAAAGAGCTGGAA	272	
Db	3604 NNTDATTDAVAGSTMGRMATTHDAYNNTKNTNTANDARCTSTRNMGKRAATVMBBNN	3663	
Qy	273 AACAAATTAATTCGTGTGCACAGTCAGTAGCAGCCGTGTTGGGTGCACATPACAGCACT	332	
Db	3664 TRACANTSNMNCASNNRGRATGASARHANDVNTCTDAND-RAATNNVSTYATH	3722	
Qy	333 TTGTTTGCAACATPACATCAATCTAAAGTGTGTTCTTTCTCTCTTCTTAATCTTGACA	392	
Db	3723 RTYADBSARCHANDNDSTRADVNNITDVBVYHRAHTATATTRYNGSRTIHACKMSNTTTA	3782	
Qy	393 GTCRAAGGTGAAGAGTAGCATATTGAGTACTTCCCTCGATCCTGTACCCAGATTAG	452	
Db	3783 TTTATDTRADARAKATTTNYSASTRAAAGACAGNANNRGMATBBSMMNDYSKDCMTD	3842	
Qy	453 CATGATATTCAAAATGAACTGAGTGAATGGAAGCAGACATATTTGGTATGATCAGCAGC	512	
Db	3843 MMYKCKOAKCDMMTNGVCHNNRBDHDDHDDRBVMBVNDVDVBNNNVDDHDDHDDH	3902	
Qy	513 -AAAGTCTTAATTAATTATGTTATATCTTCAGTAAACCTTTGCTGACAGCTGTGAGAA	571	
Db	3903 DHDHNDHCNNBVBVNDHDDDDVYNNMGRHNNHNNHNTCDHDDHNDHTCAABDCK	3962	
Qy	572 AAGAAGATTATGATTAACACSCAGATGAAATTAATCATTAAGTAGCCACCCGTGTAC	631	
Db	3963 MVVVVVVVVVVVVVVVVVVCHKHHDDHNGHMCNCCNNKMDHAAANRMACNNHHNANG	4022	
Qy	632 TGCTGTGATTAAGCAATTCATTTAGCATTTAGCATTTATTTTGAAGCTAATAGACA	691	
Db	4023 CKHADATAPAGRANTHADRTTRRAGNNBRNNMBSCDMMCKMMKMYKNDVBVRYT	4082	
Qy	692 GCAGATTTGGTGGCGTCCATAGAAACAGACTAATCATATATCTCTGAGTTTAAACAGAG	751	
Db	4083 GBBDRKQWNDYSMSBSHMNDYSKDCMNDYSTMAITRCVRSVAVSDMNDYSTMAITRCVR	4142	
Qy	752 ATTATGACACGCAAAATTTGCTCATGTTTCAAGTAGACATATCTTGCGGGAAGAAAGAGC	811	
Db	4143 YSAVASDNDYSCTMATRCVRSVAVSD-NDYDSTMAITRCVRSVAVSDNDYSCTMAT	4201	
Qy	812 TGAGCCAGATGTGCTCATTTTTCGACATTAATCCTTCAACATTTAAACCTGGAGTCTTAG	871	
Db	4202 RCVRSVAVASDNDYSTMAITRCVRSVAVSDRKSMMKSMGTMNNRMANSTYBA	4266	
Qy	872 GAATTCAAAACGTTGGTAAATATTCATTAGCAGACATCAACTACTGTAGAAATGAGC	931	

Db 4262 RACRRNHHBDSHBSSTWRDDCSTMATSSCDTMMNDYSHHSSMMNDYSNDNSC 4321
 Qy 932 AGAAACAGACATCTACTGTAATGGCTATATATAGAAATACGTAGACGTCTCTGAA 991
 Db 4322 WNYSDMCRSTWRDRSRDGRTHNAKCHTTDGDANBDDRTNFRYSHARRSRBTABAC 4381
 Qy 992 TTTAGACTACTCTTTAAAGAGTGAAGACAGCAATGAGAAATATCATCGCAATTCCTGTAG 1051
 Db 4382 ATNSADMTADMNSTDMWDSSTWRDDCMTNTSRCTNTCDNTMANASADGSTMMARY 4441
 Qy 1052 CTCAGCACTAGACTCGAAGGTCTTCTGAACCTGAACGAGTTCCTCAACCTACCTG 1106
 Db 4442 NRATNTNSTRATTCMANAGRRYWDCCNTSMARYNMAITNRBSCTRTCACHBS 4496
 RESULT 4
 ID ABR83946 standard; cDNA; 5204 BP.
 AC ABR83946;
 XX
 DT 14-AUG-2002 (first entry)
 DE Human cDNA differentially expressed in granulocytic cells #517.
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 OS Homo sapiens.
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 PF 03-OCT-2001; 2001WO-US030821.
 PR 03-OCT-2000; 2000US-0237189P.
 PA (GENE-) GENE LOGIC INC.
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 PT Detecting granulocyte activation by detecting differential expression of
 PT genes associated with granulocyte activation, which serves as diagnostic
 PT markers that is useful for monitoring disease states and drug toxicity.
 XX
 PS Claim 1; SEQ ID NO 517; 1149p; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing the
 CC expression level to an expression level in an unactivated GC, where
 CC differential expression of Gs is indicative of GCA. Also included are
 CC modulating (M2) GA by contacting GC with an agent that alters the
 CC expression of at least one gene in Gs; (2) screening (M3) for an agent
 CC capable of modulating GCA or an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease using the gene expression
 CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
 CC tissue, an allergic response in a subject, exposure of a subject to a
 CC pathogen or sterile inflammatory disease, by detecting the level of
 CC expression in a sample of the tissue of gene(s) from Gs, where the level
 CC of expression of the gene is indicative of inflammation; (4) treating
 CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile

CC inflammatory disease, by contacting a tissue having inflammation with an
 CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
 CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
 CC for screening an agent capable of modulating GCA preferably in an
 CC inflammation in a tissue; M4 is useful for detecting an inflammation
 CC (especially chronic) in a tissue, an allergic response in a subject,
 CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
 CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
 CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
 CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
 CC disease, ulcerative colitis, periodontal disease, also bacterial
 CC infection, viral infection, parasitic infection, protozoal infection,
 CC fungal infection and M5 is useful for treating one of the above
 CC conditions. The present sequence represents a gene differentially
 CC expressed in granulocytes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 5204 BP; 1770 A; 867 C; 836 G; 1731 T; 0 U; 0 Other;
 Query Match 2.1%; Score 49.8; DB 6; Length 5204;
 Best Local Similarity 59.2%; Pred. No. 0 26; Mismatches 9; Gaps 2;
 Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
 Qy 1516 TGAGCTTTAGCCGACCATCATCATGTAATGCTTCTCGATAAGCTTTCATTAAT 1575
 Db 971 TGAACCTTTAGCTTCCACATCAAGATGATGAAGTGGTTCAAGATAAATATATTAAT 1030
 Qy 1576 TCTC-----TTTGCAAAAGCTCTGCTACTTACCGAAAGTGGCTTACGACAGAAAGA 1627
 Db 1031 TCTCGGCCAAGACACACCTGATCTCTAGCTGCTGAGGCGTACT-CAACTGAATATCA 1089
 Qy 1628 TGGCATTTAAGCGTACTTGGAAAAATAGAAAAATAGAAACTATGAAAAATTCATGAG 1687
 Db 1090 TGGCGTTTACACACATTTGGAAGGTAGACCGAGTAAATCTATGCAACTTATGAGAA 1149
 Qy 1688 CAATGGGTAAAGCCTTACTTTTGAATGCTT 1718
 Db 1150 AATGGGTAAAGACTTATTTCTTGTGGCT 1180
 RESULT 5
 ID ADJ74588 standard; DNA; 5204 BP.
 AC ADJ74588;
 XX
 DT 06-MAY-2004 (first entry).
 DE Human intestinal fatty acid binding protein gene SEQ ID NO:12.
 KW de; gene; human; intestinal fatty acid binding protein; reterososis;
 KW coronary angioplasty; balloon coronary angioplasty;
 KW stent coronary angioplasty.
 OS Homo sapiens.
 PN WO2004015104-A1.
 XX
 PD 19-FEB-2004.
 PF 20-MAR-2003; 2003WO-JP003478.
 PR 09-AUG-2002; 2002JP-00233041.
 PA (NAGO-) NAGOYA IND SCI RES INST.
 PI (GIFU-) GIFU INT INST BIOTECHNOLOGY.
 XX
 DR Yamada Y, Yokota M;
 WPI; 2004-180672/17.
 XX

PT Analysis of specific gene polymorphisms in clinical nucleic acid sample
PT for prediction of risk of restenosis after balloon or stent coronary
PT angioplasty.
PS Disclosure; SEQ ID NO 12; 164bp; Japanese.
XX
XX
XX The invention relates to a novel method for predicting the risk of
CC restenosis after coronary angioplasty comprising analysing specific gene
CC polymorphisms in a clinical nucleic acid sample. The method is useful for
CC the diagnosis of the genetic risk of restenosis following balloon or
CC stent coronary angioplasty. The method has high accuracy and high
CC estimation ratio. The present sequence is used in the exemplification of
CC the invention.
XX
SQ Sequence 5204 BP; 1770 A; 867 C; 836 G; 1731 T; 0 U; 0 Other;
Query Match 2.1%; Score 49.8; DB 12; Length 5204;
Best Local Similarity 59.2%; Pred. No. 0.26;
Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
QY 1516 TGAGCTTTAGCCAGCAGCATCATCATGTAATGCTTCTGTAAGCCTGTCATTAAT 1575
DB 971 TGAACCTTTAGCTTCCACATCATCATGTAATGCTTCTGTAAGCCTGTCATTAAT 1030
QY 1576 TCTC-----TTTGAAAGCTCTGCTACTTACCAAGCTGCTTACAGAGAAAGA 1627
DB 1031 TCTCGCCCAAGAGACAGACCTGAAATCTTACTGCTTACAGAGCTGAACT -CAACTGAATATCA 1089
QY 1628 TGGCATTTTAAGCGTACTTGGAAATATAGAAATATGAAATCTGTAATTTATGGAAG 1687
DB 1090 TGGCGTTTACACAGCACTTGAAGAGTAGACCGAGTAGAAATATGACCAAGTTTATGGA 1149
QY 1688 CAATGGGTAAAGCCTTACTTTTGAATGCT 1718
DB 1150 AAATGGGTAAAGACTTATTCTTTGTGCT 1180
RESULT 6
ADO78215
ID ADO78215 standard; DNA; 5204 BP.
XX
XX ADO78215;
DT 29-JUL-2004 (first entry)
XX
DE Human fatty acid binding protein 2 Alu repeat region DNA with SNP Seq 9.
XX
XX human; db; single nucleotide polymorphism; SNP; hereditary risk;
KM coronary spasms; tumour necrosis factor alpha;
KM NADH/NADPH oxidase p22 phox gene; apolipoprotein E; apolipoprotein C-II;,
KM streptolysin 1; interleukin 6;
KM endothelium type nitrogen monoxide synthetase;
KM fatty acid binding protein 2; cardiac sudden death;
KM variant angina pectoris; myocardial infarction.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT variation /tag= a
FT /replace(2445,A)
FT /standard_name= "Single nucleotide polymorphism"
XX
XX JF2004113093-A.
XX
XX 15-APR-2004.
XX
XX 25-SEP-2002; 2002JP-00280031.
XX
XX PF 25-SEP-2002; 2002JP-00280031.
XX
XX PR 25-SEP-2002; 2002JP-00280031.
XX
XX PA (NAGO-) ZH NAGoya SANGYO KAGAKU KENKYUSHO.
XX (GIFU-) ZH GIFUKEN KOKUSAI BIO KENYUO.
XX

DR WPI; 2004-360127/34.
XX
XX Detecting genotype in nucleic acid sample useful for diagnosing risk of
PT coronary spasms, by analyzing polymorphisms in tumor necrosis factor alpha
PT gene, NADH/NADPH oxidase p22 phox gene and apolipoprotein E gene, in
PT nucleic acid sample.
PS Disclosure; SEQ ID NO 9; 599p; Japanese.
XX
XX
XX This invention relates to a novel method for detecting a hereditary risk
CC for a coronary spasms. Specifically, it refers to the identification of
CC two or more polymorphisms and determining the genotype of the nucleic
CC acid sample in order to calculate the risk of the patient to a coronary
CC spasms. The present invention describes a risk analysis that comprises
CC identifying two or more polymorphisms occurring at any position taken
CC from -863 of the tumor necrosis factor alpha gene, 245 of the NADH/NADPH
CC oxidase p22 phox gene, -219 of the apolipoprotein E gene, -482 of the
CC apolipoprotein C-III gene, -1171 of the streptolysin 1 gene, 4070 of the
CC apolipoprotein E gene, -634 of the interleukin 6 gene, -786 of the
CC endothelium type nitrogen monoxide synthetase gene or position 2445 of
CC the fatty acid binding protein 2 gene. Accordingly, the method enables
CC prevention of cardiac sudden death resulting from variant angina
CC pectoris, coronary spasms and myocardial infarction. This polymorphic
CC is the human fatty acid binding protein 2 gene Alu repeat region
CC containing a G2445A SNP in the sequence given in the invention. NOTE:
CC This sequence is referred to as Seqid 9 in the sequence listing, but
CC Seqid 8 also refers to a primer given in the disclosure of the
CC specification.
XX
SQ Sequence 5204 BP; 1770 A; 867 C; 836 G; 1731 T; 0 U; 0 Other;
Query Match 2.1%; Score 49.8; DB 12; Length 5204;
Best Local Similarity 59.2%; Pred. No. 0.26;
Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
QY 1516 TGAGCTTTAGCCAGCAGCATCATCATGTAATGCTTCTGTAAGCCTGTCATTAAT 1575
DB 971 TGAACCTTTAGCTTCCACATCATCATGTAATGCTTCTGTAAGCCTGTCATTAAT 1030
QY 1576 TCTC-----TTTGAAAGCTCTGCTACTTACCAAGCTGCTTACAGAGAAAGA 1627
DB 1031 TCTCGCCCAAGAGACAGACCTGAAATCTTACTGCTTACAGAGCTGAACT -CAACTGAATATCA 1089
QY 1628 TGGCATTTTAAGCGTACTTGGAAATATAGAAATATGAAATCTGTAATTTATGGAAG 1687
DB 1090 TGGCGTTTACACAGCACTTGAAGAGTAGACCGAGTAGAAATATGACCAAGTTTATGGA 1149
QY 1688 CAATGGGTAAAGCCTTACTTTTGAATGCT 1718
DB 1150 AAATGGGTAAAGACTTATTCTTTGTGCT 1180
RESULT 7
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
XX ADA71938;
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KM gene; de.
XX
XX Oryza sativa.
XX
XX OS
XX PN WO2003000898-A1.
XX
XX PD 03-JAN-2003.
XX
XX PF 22-JUN-2001; 2001WO-1B001105.
XX

CC From WIPO at ftp.wipo.int/pub/published_pct_sequences

CC intervention. The compounds that modulate the activity of a tumour

CC suppressor gene are useful in the treatment of cancer or as anti-cancer
 CC drugs. The present sequence represents a polynucleotide of the invention.
 XX
 SQ Sequence 662 BP; 217 A; 133 C; 148 G; 164 T; 0 U; 0 Other;
 Query Match 2.0%; Score 46.8; DB 10; Length 662;
 Best Local Similarity 73.2%; Pred. No. 0.68; Mismatches 22; Indels 0; Gaps 0;
 Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 1615 ACAGACAGAAAGATGCGATTAAACGCTTGGAAAATGAGAAAATGAACTATGAA 1674
 DB 28 ACAGCTGAGATCATGCGATTCAACGCGACGTTGAAAGTAAACCGAAGCAAGAACTATGAA 87
 QY 1675 AAATTCATGAGACATGCGTA 1696
 DB 88 AAGTTCATGAGAAATGCGCA 109
 RESULT 11
 AAL19953/C
 ID AAL19953 standard; cDNA; 610 BP.
 XX
 AC AAL19953;
 XX
 DT 07-DEC-2001 (first entry)
 XX
 DE Human breast cancer expressed polynucleotide 12410.
 XX
 KW Human; breast cancer; cell marker; cytostatic; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200151628-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 10-JAN-2001; 2001WO-US000798.
 XX
 PR 14-JAN-2000; 2000US-0176077P.
 PR 14-MAR-2000; 2000US-0189167P.
 PR 24-MAR-2000; 2000US-0192099P.
 PR 29-MAR-2000; 2000US-0193480P.
 PR 15-MAY-2000; 2000US-0205230P.
 PR 09-JUN-2000; 2000US-0211315P.
 PR 25-JUL-2000; 2000US-0220534P.
 XX
 PA (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lillie J, Xu Y, Wang Y, Steinmann K;
 XX
 DR WPI; 2001-451856/48.
 XX
 PT New peptide useful as a marker for the diagnosis of breast cancer.
 XX
 PS Claim 1; Page 2193; 3695pp; English.
 XX
 CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterizing treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity
 CC
 SQ Sequence 610 BP; 211 A; 93 C; 96 G; 205 T; 0 U; 5 Other;
 Query Match 2.0%; Score 46.6; DB 4; Length 610;
 Best Local Similarity 62.4%; Pred. No. 0.74; Mismatches 44; Indels 0; Gaps 0;
 Matches 73; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
 QY 2204 AGAATTCGTGGCTAGCTCCCTCAACTACAGAAAAGTCAATTAATAATGCAAACTGA 2263

DB 267 AAAAGGCTGTTTACCTTTTTCATCGCGAAAAAAGAAAAAAGAAAAA 208
 QY 2264 TGTTCATTTTGTGTTTCTCTGCTGATGTTAATGATTAATATATTTT 2320
 DB 207 AGGCGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 151
 RESULT 12
 ABL70479
 ID ABL70479 standard; DNA; 9742 BP.
 XX
 AC ABL70479;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated cell signalling DNA sequence#185.
 XX
 KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
 KW tumour; cytostatic; db.
 XX
 OS Unidentified.
 XX
 PN WO200202807-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 29-JUN-2001; 2001WO-EP007471.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-154758/20.
 XX
 PT Nucleic acid, useful for diagnosis and therapy of diseases associated
 PT with cell signaling e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with cell signaling.
 XX
 PS Claim 1; SEQ ID NO 369; 24pp + Sequence listing; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of at least
 CC 18 bases of a segment of chemically pretreated DNA of genes associated
 CC with cell signaling. The activity of the modified sequences of the
 CC invention may be described as cytostatic. The object of the invention is
 CC to provide the chemically modified DNA of genes associated with cell
 CC signaling, as well as oligonucleotides and/or PNA-oligomers for
 CC detecting cytosine methylations, as well as a method which is
 CC particularly suitable for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with cell signalling. The
 CC chemically modified DNA provided by the invention is useful for diagnosis
 CC and therapy of diseases such as solid tumours and cancer. The sequences
 CC given in records ABL70111-ABL70626 represent chemically pre-treated
 CC genomic DNA's of genes associated with cell signalling. Note: The
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office
 XX
 SQ Sequence 9742 BP; 3039 A; 88 C; 1939 G; 4676 T; 0 U; 0 Other;
 Query Match 2.0%; Score 46.6; DB 6; Length 9742;
 Best Local Similarity 47.7%; Pred. No. 1.9; Mismatches 149; Indels 0; Gaps 0;
 Matches 136; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
 QY 2051 TTGAGTCATTTTAACATGCTAGTCTTAACTTGAACCTGTAATATG 2110
 DB 6826 TTAGTATATTTTGTGAAGATATTTTAATATTTTATTTTGTAAATATA 6885
 QY 2111 GATGTAAATTAATCTCTAGTGTAGTATGTTGTTTACAGGCTAACCTGCTC 2170

Db	6886	TATGTAGATTTTTTAAATTTAAATTTGAATGAAAGTTGATTTATATAGTTAGATTTTTTTA	6945
Qy	2171	AGTGAAGAGGTGAGGAAGAAGTAAAGTCTGAGTCAAGAAATTCGGGCTAAGCTCCCTCACT	2230
Db	6946	GGTATATGTTGGTATATATATATATTTGGTTTATTTTAAATTTATATATATTTTATATTT	7005
Qy	2231	ACAGAAAAAGTCACAAATTAATAAATGCAACATGATGTTCTATTTTGGTTTCTCGCTGA	2290
Db	7006	ATTTTATTTATTTAAGATATATATGTTTATAGTGAATGTTATGTTTTTTTTTTTTTTTTT	7065
Qy	2291	TGTTAATTTGATTAATTAATTTATTTTTTTTTTTAGCGGTGAATGTCAT	2335
Db	7066	TATTTATGTTGTTGTTAATTTTTTTGTTTTTTTATTTTGAATAGGCT	7110
RESULT 13			
ABLJ32692			
ID	ABLJ32692	standard; DNA; 6222 BP.	
XX	XX	XX	XX
AC	ABLJ32692;		
DT	26-MAR-2002	(first entry)	
XX	XX	XX	XX
DE	Human immune system associated gene	SEQ ID NO: 665.	
XX	XX	XX	XX
KM	Human; immune system disease; cytosine methylation; antiaesthetic;		
KM	antiartherosclerotic; antianemic; cytosatic; noctropic;		
KM	neutropoietic; anti-HIV; anticonvulsant; ophthalmological;		
KM	antirheumatic; antiarthritic; antidiabetic; antiporiatic;		
KM	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;		
KM	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KM	neutofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;		
OS	ds.		
XX	XX	XX	XX
XX	Homo sapiens.		
OS	XX	XX	XX
PN	WO200200928-A2.		
PD	03-JAN-2002.		
XX	XX	XX	XX
PF	02-JUL-2001; 2001WO-EP007537.		
XX	XX	XX	XX
PR	30-JUN-2000; 2000DE-01032529.		
PR	01-SEP-2000; 2000DE-01043826.		
XX	XX	XX	XX
PA	(EPIG-) EPIGENOMICS AG.		
PI	Olek A, Piepenbrock C, Berlin K;		
DR	WPI; 2002-130909/17.		
XX	XX	XX	XX
PT	Nucleic acid comprising fragment of chemically modified gene, useful for		
PT	diagnosis and treatment of diseases associated with abnormal cytosine		
PT	methylation.		
XX	XX	XX	XX
PS	Claim 1; SEQ ID NO 665; 32pp + Sequence Listing; German.		
XX	XX	XX	XX
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neutofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel		
CC	diseases. The present sequence is a gene of the invention		
XX	XX	XX	XX
SEQ	Sequence 6222 BP; 1941 A; 55 C; 1168 G; 3058 T; 0 U; 0 Other;		
Query Match			
Beat Local Similarity 66.7%; Score 46.2; DB 6; Length 6222;			
Matches 66; Conservative 0; Mismatches 33; Indels 0; Gaps 0;			
Qy	1623	AAAGATGGCATTTAAACGTACTTGGAAAAATAGAAAAATGAGAACTATGAAAAATTCAT	16823

Db	4958	AATTATGGCGTTTGATGTAATTGGAAAGGTAGATCGAGCGAAATATATATAGTTAT	5017
Qy	1683	GGAAAGCAATGGGTAAAGCCTTACTTTTGGATGCTTCT	1721
Db	5018	GGAAAAATGGGTAAAGATTATTTATTTTGTGGTTAT	5056
RESULT 14			
ABN80056			
ID	ABN80056	standard; DNA; 10652 BP.	
AC	ABN80056;		
DT	15-JUN-2002	(first entry)	
DE	Human	chemically modified disease associated gene SEQ ID NO 73.	
KW	Human;	development; homeobox gene; HOX; diabetes; cancer; apoptosis;	
KW	heart disease; epilepsy; histone deacetylation; muscular dystrophy;		
KW	dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;		
KW	antidiabetic; cyostatic; anticonvulsant; de.		
XX	Homo sapiens.		
OS	Synthetic.		
XX	WO200200927-A2.		
PN	03-JAN-2002.		
PD	02-JUN-2001; 2001WO-EP007536.		
PP	30-JUN-2000; 2000DE-01032529.		
XX	01-SEP-2000; 2000DE-01043826.		
PR	(EPIG-) EPIGENOMICS AG.		
PA	Olek A, Piepenbrock C, Berlin K;		
XX	WPI; 2002-130908/17.		
DR	Novel	nucleic acid useful for diagnosis and therapy of diseases	
PT	associated with	development genes such as diabetes, comprises a sequence	
PT	of a segment of	chemically pretreated DNA of genes associated with	
PT	development.		
XX	Claim 1; SEQ ID NO 73; 27pp; English.		
PS	The invention	relates to a nucleic acid (I) comprising a sequence at	
XX	least 18 bases	in length of a segment of chemically pretreated DNA (II)	
CC	of genes	associated with development selected from 87 genes listed in the	
CC	specification	such as ACCPM, ADPM, or APD1 and comprising one of 350	
CC	sequences	(ABN79984-ABN80333) or their complements. The invention is	
CC	useful for the	diagnosis or therapy of diseases associated with	
CC	development	genes, in particular disease related to homeobox containing	
CC	genes (HOX), like	diabetes, cancer, apoptosis related diseases, syndromes	
CC	associated with	congenital heart disease, epilepsy, diseases related to	
CC	histone deacetylation,	Curranio syndrome, diseases related with the	
CC	development of the	brain and limb girdle muscular dystrophy and dwarfism.	
CC	Oligomers	specific to each of the genes are useful for detecting the	
CC	methylation	state of all CpG dinucleotides within the 350 sequences or	
CC	(II) and their	complementary sequences, as primer oligonucleotides for	
CC	the amplification	of the 350 sequences, (II) and/or their complements and	
CC	as oligomer	probes for detecting the cytosine methylation state and/or	
CC	single nucleotide	polymorphisms (SNPs). Note: The sequence data for this	
CC	patent did not	form part of the printed specification but is based on	
CC	sequence	information supplied to Derwent by the European Patent Office	
XX	Sequence	10652 BP; 2132 A; 601 C; 3186 G; 4733 T; 0 U; 0 Other;	
Query Match	1.9%;	Score 45; DB 6; Length 10652;	
Best Local Similarity	47.1%;	Fred. NO. 4.7; Mismatches	155; Indels
Matches 138;	Conservative	0; Gaps	0

OY	1835	ATTTGAAGAAACCTCTATPAGAGGGGAAATACAGAAAGAAAACATTCGTGATTTATTTGC	1894
Db	588	ATTGATTAATTAATTTAAATTAAGTAGAGAAAATGTTGAAATTTAAATTCGTAATTTTAGTTGT	647
OY	1895	ATTGCGATTAATCTTAGCATTTAGCTAATTCAGTAGAGCAATTCACAGAAATTTTAA	1954
Db	648	TTTTTAAATATAAAATTAATATATTTTTTATTTATTTATTTGATGATCAAGAAAATAAAATTTAGTT	707
OY	1955	TAGATATTATATGTAAAGAAATATTAATTTTGTATPAGAAGCTGTGGAATAAAATTTACACAAGAGCG	2014
Db	708	TATGATTTTAATTTAAATATATATGTTTTTTAAATAATATATTTTTTTTAGTTAGTTAGTTATTA	767
OY	2015	AAATTTCTCGTCTCCAGTTTTCGACAGACACATGATTTGAGTCATTTTAAACATGCTAGT	2074
Db	768	TAAATTTAATTAGATTTTTTTTGGTTAAGATATGATTTAGTGTGATATTTAAAGCGGGAGGT	827
OY	2075	GCTTACTTTAAGCTTGACAAACAGCCCTGTATATATGATGTGTAATCATTAACAT	2127
Db	828	GGTTGTTTAGATATTTTTTTTTTTTAGCGGAATTTAGATTAAATGAGATTATAT	880

XX	Result 15
XX	ADA71938/c
ID	ADA71938 standard; DNA; 2000 BP.
XX	
AC	ADA71938;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Rice gene, SEQ ID 5263.
XX	
KW	Plant; bacterial infection; fungal infection; viral infection; rice;
KW	gene; ds.
XX	
OS	Oryza sativa.
XX	
PN	MO2003000898-A1.
XX	
PD	03-JAN-2003.
XX	
PF	22-JUN-2001; 2001MO-IB001105.
XX	
PR	22-JUN-2001; 2001MO-IB001105.
XX	
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.
XX	
PI	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI	Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX	
DR	WPI; 2003-175290/17.
XX	
PT	Identifying at least one gene involved in plant resistance or response to
PT	pathogenic infection for conferring resistance or tolerance to a plant to
PT	bacterial, fungal or viral infection by determining or detecting plant
PT	gene expression.
XX	
PS	Claim 27; SEQ ID NO 5263; 899pp; English.
XX	
XX	The present invention relates to a method (M1) for identifying genes
CC	involved in plant resistance or response to pathogenic infection. M1
CC	comprises identifying a gene whose expression is significantly altered in
CC	the incompatible interaction of plant gene expression relative to
CC	expression of the gene in an uninfected plant, in a mutant plant that
CC	does not express a gene associated with response to pathogenic infection,
CC	or in a corresponding incompatible or compatible interaction. (M1) is
CC	useful for conferring resistance to resistance or tolerance to a plant to
CC	bacterial, fungal or viral infection. The present sequence was used to
XX	illustrate the invention.
XX	
SQ	Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
XX	
Query Match	1.9%; Score 44.4; DB 8; Length 2000;

```

Best Local Similarity 11.1%; Pred. No. 3.7;
Matches 47; Conservative 195; Mismatches 181; Indels 1; Gaps 1;
OY 1108 GGAGTTCAAGTGATCCTTCATCTCATGCTTAATATATGTGAGTAGAAATGATTTCCACCA 1167
Db 1108 ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1023 GGCTCTKMMRYTYSMSMTYAMMKRYTKYMTAYSSTWKKYWAYKMRAYAMRSRKTWCT 964
OY 1168 AATTGAAGATGACAAACAGAGATTTGTGTTTATCTGTTGGGAATAATAGCTTTTCACCA 1227
Db 963 GGGKMAATYCGTKMAAAGMRWRMRAMCMYCCMKMKWNTSCMMWRYMTWTSWYMMWGAM 904
OY 1228 GTTGATATAAGACCCCTCCACAGATATAAAGTCCATATGACCAACAAGAA-AATGTCATAC 1286
Db 903 RYATYAARRRRMYTKMSMRMYMTMTKMATYMTCMCAKMYAATGATATMMRRYMYTY 844
OY 1287 ATTCTCTTAGCTCTCATTATTAATTTTCATTAGATACCGGTTTTTTTACTACACTCAATA 1346
Db 843 CYAATCKACKCKYKMAMTKMTTACAMPATSWRMRPAMAGMKWRKXMRBRAYWMMRWMR 784
OY 1347 AGATGAACAGAGATAATGGGTTAGTGACTGTTTATAAAGAGATAATTAAGATACATATC 1406
Db 783 GMAAWMSRFRKWKKKYATRYMCMWMTWMSRRKYSRYMMSGMRMRMSARISRY 724
OY 1407 ATCATTTGAGCAATTAAGAGAGAGATTCAGCAACAGTGTGCTTAAACAAGTGAAAA 1466
Db 723 KCAATTXASSARWTKRAKRSRYRYRRKRYMKRGKTTYRYRYWRSRMRTRAMSKRRKXAGA 664
OY 1467 CAAAGTTAACTAAAGTACCCCTCTCTTGAACAAGATCAATGCAAGTGTAGGCTTAC 1526
Db 663 SMKSCMMYWGABRSMYSKYSCSAKCKCKTRMYTSSYMSGMYSGMYSSYKMSMTSJKSY 604
OY 1527 CAGC 1530
Db 603 MGRM 600

```

Search completed: November 20, 2005, 21:16:57
Job time : 1199 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2005, 19:03:25 ; Search time 6960 Seconds

(Without alignments)
13021.699 Million cell updates/sec

Title: US-10-099-663-1

Perfect score: 2381

Sequence: 1 agctctctgcgcagaaag.....atctgaagctcactatcag 2381

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	752.6	31.6	1473	8	CC250935	CH261-17C
2	99	4.2	598	7	CF250818	esa017_eo
3	95	4.0	595	7	CF250947	esa019_a0
4	85	3.6	705	5	BU355106	603474288
5	84	3.5	759	5	BU265304	603508640
6	81	3.4	829	5	BU296697	603741965
7	79.4	3.3	885	5	BU123280	603003274
8	79	3.3	814	5	BU234099	603792609
9	78.8	3.3	1080	5	BU123336	603149475
10	75.8	3.2	426	6	CD739695	40280311
11	74	3.1	576	6	CD739141	40264811
12	70.4	3.0	480	6	CD733346	40451321
C 13	62.4	2.6	1101	9	CNS00395	AL063921 Drosophila
14	61.4	2.6	885	5	BU123717	603147391
15	60.4	2.5	443	5	BX276255	BX276255
16	57.2	2.4	683	7	CN793516	4128553 B
17	55.8	2.3	657	7	CK945920	4070423 B
18	54.2	2.3	572	7	CK834687	4059069 B
19	54.2	2.3	673	7	CK947902	4072726 B
20	54.2	2.3	666	7	CK957350	4097834 B
21	54.2	2.3	708	7	CK960577	4101917 B
22	54	2.3	481	4	BM430667	BM430667
23	54	2.3	489	6	CB224035	1JB10F5
24	54	2.3	497	4	BM430956	1DU07F03

25	54	2.3	548	4	BM432416	BM432416 1JB14B.8
26	54	2.3	568	4	BM431104	BM431104 1DU09G09
27	54	2.3	671	7	CK980160	CK980160 4111791 B
28	54	2.3	679	7	CK947198	CK947198 4071820 B
29	54	2.3	682	7	CK948900	CK948900 4074042 B
C 30	54	2.3	684	7	CK947048	CK947048 4071436 B
31	54	2.3	702	7	CK944155	CK944155 4068182 B
32	54	2.3	720	7	CK946512	CK946512 4071131 B
33	53	2.2	372	4	BM430450	BM430450 1DU034E2
34	53	2.2	414	4	BM433066	BM433066 1JB13E2.a
35	53	2.2	416	4	BM432621	BM432621 1JB11H7
36	53	2.2	546	7	CN098758	CN098758 EC2CA15A
C 37	53	2.2	576	7	CN098757	CN098757 EC2CA15A
38	52.6	2.2	660	7	CK954372	CK954372 4094190 B
39	52.6	2.2	1101	9	CNS00395	AL063921 Drosophila
40	52.4	2.2	503	4	BM432863	BM432863 1JB17B11
41	52	2.2	339	4	BM430492	BM430492 1DU035B12
42	52	2.2	371	4	BM432453	BM432453 1JB14F5.a
43	52	2.2	371	4	BM433013	BM433013 1JB18H11
44	52	2.2	415	4	BM430284	BM430284 1DU032C9
45	52	2.2	453	4	BM431869	BM431869 1JB13B10

ALIGNMENTS

RESULT 1
LOCUS CC250935/c 1473 bp. DNA linear GSS 13-MAY-2003
DEFINITION CH261-17C7 Sp6.1 CH261 Gallus gallus genomic clone CH261-17C7,
genomic survey sequence.
ACCESSION CC250935
VERSION CC250935.1 GI.30587685
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1473)
Warrren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
CONTACT: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submis@wustl.edu
Insert Length: 18200 Std Error: 0.00
Seq primer: Sp6 ATTGAGTGCACCTATAG
Class: BAC ends
High quality sequence start: 239
High quality sequence stop: 912.
Location/Qualifiers
1..1473
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-17C7"
/sex="female"
/cell_line="WUCD001, inbred 256"
/note="Vector: pTARBA2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - For library and clone
ordering information: http://www.chnori.org/bacpac"

ORIGIN
Query Match 31.6%; Score 752.6; DB 8; Length 1473;
Best Local Similarity 98.3%; Pred. No. 4.3e-168;
Matches 793; Conservative 0; Mismatches 9; Indels 5; Gaps 3;
QY 1580 TTTCGAAGCTGCTACTTACCAAGATC-TGCTACAGACAGAAAGATGCATTAA-1637

```

Db      1042 TTGCAGAAAAGTTGCTTCTTACAGAAAATCTTCCCTACAGACAGAAAGATGGCATTTAAC 983
      |||
Qy      1638 CCGTACTTGAAAATATAGAGAAAATGAGAACTATGAAAAATTCATGAAAGCAATGGGTAA 1697
      |||
Db      982 CCGTACTTGAAAATATAGAGAAAATGAGAACTATGAAAAATTCATGAAAGCAATGGGTAA 923
      |||
Qy      1698 GCGTTACTTTTGAATGCTTCTTAAGACAGATACCACTACGGGGGAATACAACTTA 1757
      |||
Db      922 GCGTTACTTTTGAATGCTTCTTAAGACAGATACCACTACGGGGGAATACAACTTA 863
      |||
Qy      1758 AGCTGTTCAAGAACTACATCTGGCTACACCTGCTCTTGTGCTGCTATTTTGGCCCTTG 1817
      |||
Db      862 AGCTGTTCAAGAACTACATCTGGCTACACCTGCTCTTGTGCTGCTATTTTGGCCCTTG 803
      |||
Qy      1818 CACATTGCGCTGACCTTATTTTGAAGAAAGCTCTATAGAGGGGAATACAAAGAAAGAAAAC 1877
      |||
Db      802 CACATTGCGCTGACCTTATTTTGAAGAAAGCTCTATAGAGGGGAATACAAAGAAAGAAAAC 743
      |||
Qy      1878 ATTCTGATTTTATTTGCAATTTGCAATATCTTAATGCAATTTAGCTAATTCAGATAGAGGCAT 1937
      |||
Db      742 ATTCTGATTTTATTTGCAATTTGCAATATCTTAATGCAATTTAGCTAATTCAGATAGAGGCAT 683
      |||
Qy      1938 TCCAGAGAAATTTAATAGAAATTTATATGTAAGAAATTTATTTGATTAAGACTGTTTGA 1997
      |||
Db      682 TCCAGAGAAATTTAATAGAAATTTATATGTAAGAAATTTATTTGATTAAGACTGTTTGA 623
      |||
Qy      1998 AAAATTTACACAGAGGAAATTTGCTGCTCTCCAGTTTTCAGACACACATGATTTGAGT 2057
      |||
Db      622 AAAATTTACACAGAGGAAATTTGCTGCTCTCCAGTTTTCAGACACACATGATTTGAGT 563
      |||
Qy      2058 CATTTTAATGATGCTAGTCTTACTTATAGCTGTGTAACAACTGCTGTATATATGATGATTA 2117
      |||
Db      562 CATTTTAATGATGCTAGTCTTACTTATAGCTGTGTAACAACTGCTGTATATATGATGATTA 503
      |||
Qy      2118 ACATACTATCTAGTGTGATGATGATTTGATTTTACAGAGCTGTAACCTGCTCAGTGA 2177
      |||
Db      502 ACATACTATCTAGTGTGATGATGATTTGATTTTACAGAGCTGTAACCTGCTCAGTGA 443
      |||
Qy      2178 GGTGAGAAAGATTAAGACTGTAAGTCAAGAAATTTGCGGCTTAAGTCCCTCACTACAGAA 2237
      |||
Db      442 GGTGAGAAAGATTAAGACTGTAAGTCAAGAAATTTGCGGCTTAAGTCCCTCACTACAGAA 383
      |||
Qy      2238 AAGTCAATTAATAATGCAAGCAAGTCTATTTTGTCTCTGCTGATGATTAAT 2297
      |||
Db      382 AAGTCAATTAATAATGCAAGCAAGTCTATTTTGTCTCTGCTGATGATTAAT 323
      |||
Qy      2298 TGATTAATTAATAATTTTGTCTCTGCTGATGATGATTAAT 2354
      |||
Db      322 TGATTAATTAATAATTTTGTCTCTGCTGATGATGATTAAT 263
      |||
Qy      2355 CACGATTAATCTGAAGCTCACTATTCAG 2381
      |||
Db      262 CACGATTAATCTGAAGCTCACTATTCAG 236
      |||
RESULT 2
LOCUS   CF250818 598 bp mRNA linear EST 07-AUG-2003
DEFINITION esao19_e05 Elmeiria tenella-infected caecal tonsil Gallus gallus
CDNA, mRNA sequence.
ACCESSION CF250818
VERSION   CF250818.1 GI:33484073
KEYWORDS EST.
SOURCE   Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 598)
Witzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whitaker,C.J.,
Chausse,A.M. and Zoorob,R.
A collection of chicken ESTs from activated immune cells
TITLE
AUTHORS

```

```

JOURNAL Unpublished (2003)
COMMENT Contact: Zoorob R
UPR 1983
CNRS
7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoorob@vjf.cnrs.fr.
FEATURES
source
1..598
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone_lib="Elmeiria tenella-infected caecal tonsil"
/note="Organ: Caecal tonsil; Vector: pTribEX2"
ORIGIN
Query Match 4.0%; Score 95; DB 7; Length 598;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1600 ACCAGAACTGCTGCTTACAGACAGAAAGATGGCATTTAAGCTTATGAAAAATAGAGAA 1659
      |||
Db      3 ACCAGAACTGCTGCTTACAGACAGAAAGATGGCATTTAAGCTTATGAAAAATAGAGAA 62
      |||
Qy      1660 AATGAGAACTATGAAAAATTCATGGAAGCAATGG 1694
      |||
JOURNAL Unpublished (2003)
COMMENT Contact: Zoorob R
UPR 1983
CNRS
7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoorob@vjf.cnrs.fr.
FEATURES
source
1..595
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone_lib="Elmeiria tenella-infected caecal tonsil"
/note="Organ: Caecal tonsil; Vector: pTribEX2"
ORIGIN
Query Match 4.0%; Score 95; DB 7; Length 595;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1600 ACCAGAACTGCTGCTTACAGACAGAAAGATGGCATTTAAGCTTATGAAAAATAGAGAA 1659
      |||
Db      3 ACCAGAACTGCTGCTTACAGACAGAAAGATGGCATTTAAGCTTATGAAAAATAGAGAA 62
      |||
Qy      1660 AATGAGAACTATGAAAAATTCATGGAAGCAATGG 1694
      |||

```

Db 63 AATGAGAACTATGAAAAATTCATGAGCAATGG 97

RESULT 4
LOCUS BU265304 705 bp mRNA linear EST 28-NOV-2002
DEFINITION 603474288F1 CSEQCHN70 Gallus gallus cDNA clone CHEST355022 5', mRNA sequence.

ACCESSION BU265304
VERSION BU265304.1 GI:25536254
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J. A Comprehensive Collection of Chicken cDNAs Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE JOURNAL MEDLINE PUBMED
22335534 12445392

COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source location/Qualifiers
1..705
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST355022"
/dev_stage="36"
/lab_host="DH10B"
/clone_1b="CSEQCHN70"
/note="Organ: hearts; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntend, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN
Query Match 3.6%; Score 85; DB 5; Length 705;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1610 TGCCTACAGACAGAAAGATGCGATTAAACGTAATTGAGAAATAGAGAAATGAGAACT 1669
DB 1 TGCCTACAGACAGAAAGATGCGATTAAACGTAATTGAGAAATAGAGAAATGAGAACT 60
QY 1670 ATGAAATTCATGAGCAATGG 1694
DB 61 ATGAAATTCATGAGCAATGG 85

RESULT 5
BU265304

LOCUS BU265304 759 bp mRNA linear EST 26-NOV-2002
DEFINITION 603508640F1 CSEQCHN52 Gallus gallus cDNA clone CHEST437010 5', mRNA sequence.

ACCESSION BU265304
VERSION BU265304.1 GI:25536254
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J. A Comprehensive Collection of Chicken cDNAs Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE JOURNAL MEDLINE PUBMED
22335534 12445392

COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source location/Qualifiers
1..759
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST437010"
/dev_stage="22"
/lab_host="DH10B"
/clone_1b="CSEQCHN52"
/note="Organ: limbs; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntend, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN
Query Match 3.5%; Score 84; DB 5; Length 759;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1611 GCTTACAGACAGAAAGATGCGATTAAACGTAATTGAGAAATAGAGAAATGAGAACTA 1670
DB 1 GCTTACAGACAGAAAGATGCGATTAAACGTAATTGAGAAATAGAGAAATGAGAACTA 60
QY 1671 TGAATTCATGAGCAATGG 1694
DB 61 TGAATTCATGAGCAATGG 84

RESULT 6
BU296697
LOCUS BU296697 829 bp mRNA linear EST 27-NOV-2002
DEFINITION 603741965F1 CSEQCHN56 Gallus gallus cDNA clone CHEST64017 5', mRNA sequence.

ACCESSION BU296697
VERSION BU296697.1 GI:25746333
KEYWORDS EST.

SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
Gallus gallus (chicken)	Gallus gallus	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.	1 (bases 1 to 829)	Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.	A Comprehensive Collection of Chicken cDNAs	Curr. Biol. 12 (22), 1965-1969 (2002)	
12445392	Contact: Simon Hubbard						
Department of Biomolecular Sciences							
University of Manchester Institute of Science and Technology							
(UMIST)							
PO Box 88, Manchester, M60 1OD, UK							
Tel: 01612008930							
Fax: 01612360409							
Email: Simon.Hubbard@umist.ac.uk.							
Location/Qualifiers							
1. .829							
/organism="Gallus gallus"							
/mol_type="mRNA"							
/strain="Compton line 151"							
/db_xref="taxon:9031"							
/clone="CHEST640b17"							
/sex="Female"							
/dev_stage="adult"							
/lab_host="DH10B"							
/clone_lib="CSEQCIN56"							
/note="Organ: small intestine, Vector: pBluescript II							
KS(+); Site_1: EcoRI; Site_2: NotI; This normalized							
library was constructed from 1 million independent clones.							
cDNA synthesis was initiated using an oligo(dT) primer,							
using methylated C in the first strand synthesis reaction.							
Following this first strand reaction, double-stranded cDNA							
was blunted, ligated to NotI adapters, digested with							
EcoRI, size-selected, and cloned into the NotI and EcoRI							
compatible sites of a custom modified MCS of the							
pBluescript (KS+) vector. The library was normalized in 2							
rounds using conditions adapted from Soares et al., PNAS							
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6							
(1996): 791, except that a significantly longer							
reannealing hybridization was used."							
ORIGIN							
Query Match	3.4%; Score 81; DB 5; Length 829;						
Best Local Similarity	100.0%; Pred. No. 3.9e-08;						
Matches	81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1614 TACAGACGAAAGATGCGATTTAAGGTAAGTGAAGAAATAGAGAAATGAGAACTATGA 1673						
Db	11 TACGACAGAAAGATGCGATTTAAGGTAAGTGAAGAAATAGAGAAATGAGAACTATGA 70						
QY	1674 AAAATTGATGGAAGCAATGGG 1694						
Db	71 AAAATTGATGGAAGCAATGGG 91						
RESULT 7							
LOCUS	BUI23280 885 bp mRNA linear EST 25-NOV-2002						
DEFINITION	603003774F1 CSEQCIN18 Gallus gallus cDNA clone CHEST14c20 5', mRNA						
ACCESSION	BUI23280						
VERSION	BUI23280.1						
KEYWORDS	EST.						
SOURCE	Gallus gallus (chicken)						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.						

REFERENCE	1 (bases 1 to 885)	
AUTHORS	Boatman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
TITLE	A Comprehensive Collection of Chicken CDNAs	
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	
MEDLINE	22335534	
PUBMED	12445392	
COMMENT	Contact: Simon Hubbard Department of Biomolecular Sciences University of Manchester Institute of Science and Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK Tel: 01612008930 Fax: 016123560409 Email: Simon.Hubbard@umist.ac.uk. Location/Qualifiers 1. 885	
FEATURES		
source	/organism="Gallus gallus" /mol_type="mRNA" /strain="Compton Line 151" /db_xref="taxon:9031" /clone="CHEST14c20" /sex="Female" /dev_stage="adult" /lab_host="DH10B" /clone_id="CSBQCHL18" /note="Organ: small intestine; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+) [stratagene] vector to accommodate cDNA produced with the T-timed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspI and BamHI sites [5'ggcgcgcgtgcagcccgcatccgcgaataaag] [5'aatctcttttttggatccggggctgcagc]"	
ORIGIN		
Query Match	3.3%;	Score 79.4; DB 5; Length 885;
Best Local Similarity	98.8%;	Pred. No. 9.6e-08;
Matches	80; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1614 TACAGACAGAAAGATGCGATTTAACGGTACTTGGAAAAATAGAGAAAATGAAACTATGA	1673TATTTT
Db	1 TTTCAGACAGAAAGATGCGATTTAACGGTACTTGGAAAAATAGAGAAAATGAAACTATGA	60
QY	1674 AAAATTCAATGGAACCAATGGG	1694
Db	61 AAAATTCAATGGAACCAATGGG	81
RESULT 8		
BU234099	814 bp	mRNA linear EST 26-NOV-2002
DEFINITION	603792609F1 CSBQCHN24 Gallus gallus cDNA clone CHEST758b4 5', mRNA	
LOCUS	BU234099	sequence.
ACCESSION	BU234099	
VERSION	BU234099.1	GI:25478348
KEYWORDS	EST.	
SOURCE	Gallus gallus (chicken)	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 814)	
REFERENCE	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.	
AUTHORS	A Comprehensive Collection of Chicken CDNAs	
TITLE	Curr. Biol. 12 (22), 1965-1969 (2002)	
JOURNAL	22335534	
MEDLINE	12445392	
PUBMED	Contact: Simon Hubbard	
COMMENT		

Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

Location/Qualifiers

1..814

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="White Leghorn, H1sex"

/db_xref="taxon:9031"

/clone="CHEST758b4"

/dev_stage="22"

/lab_host="DH10B"

/clone_1ib="CSBOCHN24"

/note="Organ: heads; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS 1994) 91: 9228-9232 and Bonaldi et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 3.3%; Score 79; DB 5; Length 814;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 CAGACGAAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAGCACTATGAAA 1675

DB 1 CAGACGAAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAGCACTATGAAA 60

QY 1676 AATTCTGAGAGCAATGCG 1684

DB 61 AATTCTGAGAGCAATGCG 79

RESULT 9

BUI23336

LOCUS BUI23336 1080 bp mRNA linear EST 25-NOV-2002
DEFINITION 603189475F1 CSBQCHL18 Gallus gallus cDNA clone CHEST15319 5', mRNA sequence.

ACCESSION BUI23336

VERSION BUI23336.1

KEYWORDS GI:25333972

SOURCE EST

ORGANISM Gallus gallus (chicken)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus

1 (bases 1 to 1080)

REFERENCE Boardman, P.E., Sanz-Eguero, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

JOURNAL MEDLINE 2233534

PUBMED 12445392

COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

1..1080

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="Compton line 151"

/db_xref="taxon:9031"

/clone="CHEST15319"

/sex="Female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_1ib="CSBQCHL18"

/note="Organ: small intestine; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; Modification of pBluescript II KS(+) (Stratagene) vector to accommodate cDNA produced with the T-primed protocol (Construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Ry, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspI and BamHI sites [5'ggcgcgtcagcccgccgacccgcaaaaag] [5'aattcttttcggatccggcgctcgacgc]"

ORIGIN

Query Match 3.3%; Score 78.8; DB 5; Length 1080;
Best Local Similarity 97.6%; Pred. No. 1.4e-07;
Matches 80; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1613 CTACAGACGAAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAGCACTATG 1672

DB 7 CAGACGAAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAGCACTATG 66

QY 1673 AAAAATTCATGAGCAATGCG 1694

DB 67 AAAAATTCATGAGCAATGCG 88

RESULT 10

CD739695

LOCUS CD739695 426 bp mRNA linear EST 26-JUN-2003
DEFINITION 4028031 IGH1 - Chicken Intestinal Lymphocyte Gallus gallus cDNA

ACCESSION CD739695

VERSION CD739695.1

KEYWORDS GI:32290544

SOURCE EST

ORGANISM Gallus gallus (chicken)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus

1 (bases 1 to 426)

REFERENCE Min, W., Lillehoj, H.S., Ashwell, C.M., Matukumalli, L.K., van Tassel, C. and Han, J.Y.

Chicken Intestinal Lymphocyte EST database as a resource for the analysis of mucosal immune function

Unpublished (2003)

CONTACT: Hyun S. Lillehoj

Animal Parasite Diseases Laboratory

Animal and Natural Resources Institute, USDA

Bldg. 1043, BARC-East, Beltsville, MD 20705, USA

Tel: 3015048771

Fax: 3015045103

Email: lillehoj@anri.barc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

0.000925 using options -trim_alt - -trim_fasta. Vector identified

by cross_match using options -mismatch 12 -mnscore 12

Plate: 20 row: B column: 02

Seq primer: ATTAGGTGACACTATAG

High quality sequence scop: 426.

Location/Qualifiers

1..426

/organism="Gallus gallus"

/mol_type="mRNA"
/strain="white leghorn SC"
/db_xref="taxon:9031"
/clone="IGAL_20802"
/sex="mixed"
/tissue_type="Gut"
/cell_type="Lymphocyte"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Intestine; Vector: PCMV-SPORT6; Site 1: SalI; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."

ORIGIN

Query Match 3.2%; Score 75.8; DB 6; Length 426;
Best Local Similarity 97.5%; Pred. No. 5.8e-07;
Matches 77; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1616 CAGACGAAGAATGCGATTAAAGCTTGAAGAAATAGAGAAATGAACTATGAAA 1675
|||||
DB 2 CAGACGAAGAATGCGATTAAAGCTTGAAGAAATAGAGAAATGAACTATGAAA 61
|||||

QY 1676 AATTCATGGAAGCAATGGG 1694
|||||
DB 62 AATTCATGGAAGCAATGGG 80
|||||

RESULT 11
CD7339141 576 bp mRNA linear EST 26-JUN-2003
LOCUS CD7339141
DEFINITION clone IGAL_93A12 5', mRNA sequence.
ACCESSION CD7339141 GI:32289990
VERSION CD733914.1 GI:32289990
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 576)
Min.W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van
Tassel,C. and Han,J.Y.
Chicken intestinal lymphocyte EST database as a resource for the
analysis of mucosal immune function
Unpublished (2003)
Contact: Hyun S. Lillehoj
Animal Parasite Diseases Laboratory
Animal and Natural Resources Institute, USDA
Bldg.1043, BARC-East, Beltsville, MD 20705, USA
Tel: 3015048771
Fax: 3015045103
Email: hllilleho@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim_alt '' -trim_fasta. Vector identified
by cross_match using options -mismatch 12 -minscore 12
Plate: 93 row: A column: 12
Seq primer: ATTTAGGTGACACTATAG
High quality sequence stop: 576.
Location/Qualifiers
1..576
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="white leghorn SC"
/db_xref="taxon:9031"
/clone="IGAL_93A12"
/sex="mixed"
/tissue_type="Gut"
/cell_type="Lymphocyte"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_1lb="IGAL - Chicken Intestinal Lymphocyte"

/note="Organ: Intestine; Vector: PCMV-SPORT6; Site 1: SalI; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."

ORIGIN

Query Match 3.1%; Score 74; DB 6; Length 576;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1621 AGAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAACTATGAAAATTC 1680
|||||
DB 1 AGAAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAACTATGAAAATTC 60
|||||

QY 1681 ATGGAAGCAATGGG 1694
|||||
DB 61 ATGGAAGCAATGGG 74
|||||

RESULT 12
CD733346 480 bp mRNA linear EST 26-JUN-2003
LOCUS CD733346
DEFINITION clone IGAL_41B15 5', mRNA sequence.
ACCESSION CD733346 GI:32284195
VERSION CD733346.1 GI:32284195
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 480)
Min.W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van
Tassel,C. and Han,J.Y.
Chicken intestinal lymphocyte EST database as a resource for the
analysis of mucosal immune function
Unpublished (2003)
Contact: Hyun S. Lillehoj
Animal Parasite Diseases Laboratory
Animal and Natural Resources Institute, USDA
Bldg.1043, BARC-East, Beltsville, MD 20705, USA
Tel: 3015048771
Fax: 3015045103
Email: hllilleho@anri.barc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
0.000925 using options -trim_alt '' -trim_fasta. Vector identified
by cross_match using options -mismatch 12 -minscore 12
Plate: 41 row: B column: 15
Seq primer: ATTTAGGTGACACTATAG
High quality sequence stop: 480.
Location/Qualifiers
1..480
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="white leghorn SC"
/db_xref="taxon:9031"
/clone="IGAL_41B15"
/sex="mixed"
/tissue_type="Gut"
/cell_type="Lymphocyte"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_1lb="IGAL - Chicken Intestinal Lymphocyte"
/note="Organ: Intestine; Vector: PCMV-SPORT6; Site 1: SalI; Site 2: NotI; Normalized library from chicken gut infected with coccidia duodenum and middle gut."

ORIGIN

Query Match 3.0%; Score 70.4; DB 6; Length 480;
Best Local Similarity 98.6%; Pred. No. 1.2e-05;
Matches 71; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1623 AAGATGCGATTAAAGCTTGAAGAAATAGAGAAATGAACTATGAAAATTCAT 1682


```

Db      1  AAAGATGCAATTTAAAGGACCTTGGAAATAGAGAAATAGACACTATGAAAAATTCAT 60
Qy      1683  GGAGCAATGG 1694
Db      61  GGAGCAATGG 72

RESULT 13
LOCUS   CNS0039G/c 1101 bp  DNA  linear  GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCL-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL063921
VERSION   AL063921.1  GI:4941778
KEYWORDS  GSS.
SOURCE    Drosophila melanogaster (fruit fly)
ORGANISM  Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osogawa and
          Aaron Mamoser in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCL-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
  source
    1..1101
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"
    /clone_1ib="BACR08K10"
    /clone_1ib="RPCL-98"
    /note="end : TET3"

ORIGIN
Query Match      2.6%; Score 62.4; DB 9; Length 1101;
Best Local Similarity 17.6%; Pred. No. 0.0011;
Matches 89; Conservative 225; Mismatches 191; Indels 1; Gaps 1;

Qy      1842  AAGAAGCTTATAGAGGGAATACAGAGAAAGAAATTCATCTGATTTTATTTGATTCGA 1901
Db      985  DADGAGRRGRGRRKKDKRKGDDDKGGKKKAKAKAKATKWDMDMDMDMDMDA 926

Qy      1902  TTAATCTTATGCAATTTAGCTAATTCAGTAGAGGCAATTCAGAGAAATTTAATAGATT 1961
Db      925  KDRKADDDDDGADDDGKDDDDTDTGKDDDDDKKDDMDWKAKGTGDAATWAAAYD 866

Qy      1962  ATATGTAAGGAATATTTATTTGATAGCTGTTGAAGAAATTAACAAGAGGAATTCG 2021
Db      865  WWWGAGADDDWMDAADDDWMDADDDWMDAMWMDAMWMDAMGATADRRMDGDAAGRGGA 806

Qy      2022  TGGTCTCCAGTTTTCGACACACATGATTTGAGTCAATTTTAACTAGTGTACTT 2081
Db      805  RRRBRKRAADDKRDADDDDAATWTTTTRDTRDDMDKWKTDITRWAADRTWDRDD 746

```

```

Qy      2082  TTAAGCTTGACAAAGCTGCGTGAATATGATTAACATTAATCTCTAGTTGATACT 2141
Db      745  DDDRAGTGRKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWK 666

Qy      2142  AGTTTGTATTAAGCTGACGCTGACCTGCTGAGTGAAGTAGAGTAGAGTCTGAG 2201
Db      685  RRRDDATWDRDMDADAAWMTTDTDTDDDKRRRKRKRRRRRTTABAAMDWWTWK 626

Qy      2202  TCAGAAATTCGGGCTTAAAGTCCCTCACTACAGAAAAAGTCACAAATAAATGCAACAT 2261
Db      625  -AMDWAKMDWKTRADRRWADWTDARADRDWAKABWRARBRARABRARBRWTK 567

Qy      2262  GAGTCTCATTTTGTCTTCTCTGCTGATGATTAATGATATATATATATTTTATTTT 2321
Db      566  GKTTATWTTTAAAPAAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 507

Qy      2322  GCGTGAATGTGATGAGAAAGAAAGTT 2347
Db      506  WMTATWAAWTAAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 481

RESULT 14
LOCUS   BU123717 885 bp  mRNA  linear  EST 25-NOV-2002
DEFINITION 603147391F1 CSECHL18 Gallus gallus cDNA clone CHEST148F24 5', mRNA
            sequence.
ACCESSION BU123717
VERSION   BU123717.1  GI:25334366
KEYWORDS  EST.
SOURCE    Gallus gallus (chicken)
ORGANISM  Gallus gallus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
          Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 885)
AUTHORS   Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
          Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
          A Comprehensive Collection of Chicken cDNAs
          Curr. Biol. 12 (22), 1965-1969 (2002)
JOURNAL    2235534
MEDLINE    12445392
PUBMED     Contact: Simon Hubbard
          Department of Biomedical Sciences
          University of Manchester Institute of Science and Technology
          (UMIST)
          PO Box 88, Manchester, M60 1QD, UK
          Tel: 01612008930
          Fax: 01612360409
          Email: Simon.Hubbard@umist.ac.uk.
FEATURES
  source
    1..885
    /organism="Gallus gallus"
    /mol_type="mRNA"
    /strain="Compton line 151"
    /db_xref="taxon:9031"
    /clone_1ib="CHEST148F24"
    /sex="Female"
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Organ: small intestine; Vector: pBluescript II
    KS(+); Site 1: EcoRI; Site 2: NotI; Modification of
    pBluescript II KS(+). [Stratagene] Vector to accommodate
    cDNA produced with the T-primed protocol (Construction of
    uni-directionally cloned cDNA libraries from messenger RNA
    U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
    NotI and EcoRI. Ligate in double stranded adaptor
    containing BglI and BamHI sites
    [5'ggcgagcgagccgcccgcgacccgaaagaag]
    [5'atctcttttttcgagtcgcggcgacgcg]"
ORIGIN

```

Query Match 2.6%; Score 61.4; DB 5; Length 885;
 Best Local Similarity 96.1%; Pred. No. 0.0019;
 Matches 74; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 1618 GACAGAAAGTGCATTTAAACGTACTTGAAAATAGAGAAAATGAGAACTATGAAAA 1677
 |||
 DB 8 GAGAGAAAGATGCATTTAAACGTACTTGAAAAT-AGAAAAATGAGAACTATGAAAA 65
 |||

QY 1678 TTCATGGAAGCAATGGG 1694
 |||

DB 66 TTCATGGAAGCAATGGG 82

DB 31 TGGGCTGATGTGATGAAAAAGTTAGAGCCACGATTAATCTGAAGCTCACTATTTC 90
 QY 2380 AG 2381
 ||
 DB 91 AG 92

Search completed: November 21, 2005, 01:59:31
 Job time : 6965 secs

RESULT 15

LOCUS BX276255 443 bp mRNA linear EST 15-JUL-2004
 DEFINITION BX276255 AGENAE Gallus gallus multi-tissues normalized library
 (gcag) Gallus gallus CDNA clone gcag0008c.c.24 5prim, mRNA
 sequence.

ACCESSION BX276255
 VERSION BX276255
 KEYWORDS BX276255.1 GI:28598746
 EST.

SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 443)
 Herault,F., Le Meuth-Metzinger,V., Desert,C., Retout,E., Pluiml,F.,
 Klopp,C. and Douaite,M.
 Construction and primary characterization of chicken normalized
 multi-tissue cDNA libraries
 Unpublished (2003)
 Contact: Douaite M
 INRA, UMR INRA-ENSAR Genetique Animale
 65, rue de Saint-Brieuc, RENNES cedex, 35042, FRANCE
 Tel: +33 (0) 2.23.48.54.63
 Fax: +33 (0) 2.23.48.54.70
 Email: Madeleine.Douaite@renn.inra.fr
 Sequence cleaned of vector, adaptor and repetitions. Contact us
 at sigenastupport@jouy.inra.fr to obtain the chromatogram of this
 sequence.
 Plate: 0008 row: c column: 24
 Seq primer: M13R.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 Location/Qualifiers
 1..443
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 /clone="gcag0008c.c.24"
 /tissue_type="multi-tissues"
 /dev_stage="from embryos to adults"
 /lab_host="DH10B"
 /clone_1ib="AGENAE Gallus gallus multi-tissues normalized
 library (gcag)"
 /note="Vector: pTR73D-pac; tissues: brain, embryos,
 kidney, multi-tissues, muscle, pancreas, skin, testis,
 liver, adipose tissue, granulosa, utero-vaginal gland,
 oviduct, small follicle, ovary, hypothalamus, pituitary
 gland, ileum, jejunum, caecum, duodenum, spleen,
 fabricius gland, bone marrow, thymus, hematopoietic
 progenitor cells. Clone distribution : AGENAE Resource
 centre, Francois PIUMI, Francois PIUMI, INRA, CEA
 Radiobiologie et Etude du genome (LREG), Domaine de
 Vilvert, 78352, Jouy-en-Josas cedex, FRANCE"

ORIGIN

Query Match 2.5%; Score 60.4; DB 5; Length 443;
 Best Local Similarity 98.4%; Pred. No. 0.0027;
 Matches 61; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2320 TAGGCGTGAATGTGATGAAAAAGTTAGAGCCACGATTAATCTGAAGCTCACTATTTC 2379
 |

1123 CCITCAITCICHA
QY

		: OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. <i>FAB1</i> promoter			
		US-09-806-708B-22			
Query Match	2.7%;	Score 64.2;	DB 4;	Length 1141;	
Best Local Similarity	12.7%;	Pred. No. 4.3e-07;			
Matches 136;	Conservative 379;	Mismatches 550;	Indels 10;	Gaps 3;	
Qy	830	TTTTGCATTAATCCTTCAACATTTAAACCTGGGATCTAGTGAATCAACACGTTGGG	889.		
Db	1129	TTTCTKTKKANNNNNNNNNGGMDNNMDTKSANGTATWTHAKKGATWCTWYWTGTGN	1070		
Qy	890	TAAATTTCACTTAGACGACATCAATCACTGTAGAGATGACAGAAACAGACATTTCACT	949		
Db	1069	RRRCMTYAMRTYTHSNMNSCATBEMWTKKWTATKTRTAMVYAMCWMRNNNNMCATN	1010		
Qy	950	GAATGGGCTATATAATATAGAAATACGTAGAGAGTGTCTGAAATTTAGACTACTTATAA	1009		
Db	1009	GYAKSCATNNMAYATTBMAAAYAAAKWABAGNNEMRYGAAAGNKKGMAAATMGBMW	950		
Qy	1010	GATGAGGACACGAAATGAGAAATATCATCGCAATTTCTGTACTGACGACTAGATCGAA	1069		
Db	949	DTAGKACNNNNNNWTTDVERBMAKAKNNNNNNNAIYWPACYNRAATNNKATTHMKWTHGH	890		
Qy	1070	GGTTTCTGAAACGTAGACCGAG-----TTTCCCAACATCACTGTGAGATGTCAGTGAT	1122		
Db	889	SKRTTGHRTCTRTKNNNNNNNAATVYVYHHAARBMAMWTRTNNNNNNNNAACRAT	830		
Qy	1123	CTTCACTTCGATGCTTATTAATGAGGAAATAGATTCTCCAAATTAGATGACAA	1182		

[illegible]

US-09-806-708B-22

```
Query Match      2.4%; Score 57.8; DB 4; Length 1141;
Best Local Similarity 10.9%; Fred. No. 2.3e-05;
Matches 82; Conservative 263; Mismatches 407; Indels 0; Gaps 0;

Oy 1565 TGTCATTAATTCCTCTTGGCAAAGCTGCTGACTTACAGAGAGTGTGCTTACAGACAGAA 1624
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 DRDMSBKRMVYGMGBWKKMSYDYVYVYVWDMDCRKXRYRWRVTRTGMRMNVYVAHTAIR 260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1625 AGATGGCATTAAACGGTACTTGGAATAATAGAAAAATGAGAACTATGAAAAATTCATGG 1684
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 RRYNNGWTBAAAYRWRMTVNNNNNNNAKAKCKAKYGMWRABVNSTCTTKWSKTTKVRTSC 320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1685 AAGCAATGGGTAAAGCTTACTTTTGAATGCTCTTAAACAGAGATCCACTACGGGG 1744
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 WANNCRADDAKDHMKWMSAAMGVYNNNNNNNTYKKAIRBAMBDVWMSAKMHAN 380
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1745 GAATACAAACTTAAGCTGTTCATGAACCTACCACTGGGCTAACCTGTCCTTGTGTGCTGC 1804
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 AAHSRKKWTYKRTTVMNNNNGTMMKGMAMAYKMDMDGCTVNNNNNGRTTYGWRK 440
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1805 TATTTTCCCTTGACATTTGCCCTGCACCTATTGTTAAAAAGCTATAGAGGAGATAC 1864
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 NKQWVTTYKMKANNCKMRAMDHKCTHNTTMMKKKTYWNNCYWMSMTGKSHBBAAY 500
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1865 AAGGAGAAAAAATCTTGAATTTATTTGCAATGCGATATCTTATGCAATTTAGCTAAT 1924
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 TWYMMWRVYAHANNMNDYWKACTWYKYBVCSKMNMYAAMYTKSWMYNTSRYYRMKT 560
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1925 CCAGTAGGAGCATCCAGACAGAAATTTAAATGATTTATGTAAGGAATATTATTATTGA 1984
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 NNSWRKSRDTSMGHRANNYARABHGYKMTNRWMSHTBHBRAALAHYMBMYBAKC 620
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 1985 TAAGACTGTTGAAATAATTAACAAGAGGAAATGCTGCTCCAGTTTTCAGAAACAC 2044
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 621 HCKMAYKAKKYAAGAGSGSNNNNNNNNNNNNNNNNNNNATCABDDYVYASRWYAMAAKMYRK 680
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 2045 ACATGATTTAGTCAATTTTAAACATGCTAGTCTTACTTAAAGCTGTGACAAATGCTGT 2104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 681 BAANNAYYTHANNMWGCMWNNATDTRRTYMKNNNNNNNAGWKNNNNNNNAKAAAKYVAAA 740
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 2105 AATATGATGTAAACATATACATATCTCTAGTGTGATAGTAGTGTGATTAACGCTGAAC 2164
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 741 AVKAKKRWANRANKMARGMHADAABTTDKNNNGAYTYTNNNTYRGVNTPTAAD 800
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 2165 TGCTCAGTGAAGAGGTGAGAAAGATAGACTGTGACAGATTTCTGGGCTAGCTCCC 2224
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 801 GMANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 860
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 2225 TCAACTACAAAAAGTACACATATAAAATGCAACATGATGTTCTATTTTGTCTCT 2284
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 861 WRBAYTNNNNNNRMYAGYADVAYVMSDTCDAWKKMATKNNATYTRGTARPTNNN 920
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 2285 GCTTGATCTAATGATTAATTAATTTATTTT 2316
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 NNNMTKTKYYBHAAAMNNNNNNNNKXCTAHTW 952
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-949-016-11761
; Sequence 11761, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE. METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
```


TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:


```

: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 18020
:
: LENGTH: 601
:
: TYPE: DNA
:
: ORGANISM: Human
:
: US-09-949-016-18020

```

[illegible]

RESULT 13
 US-09-949-016-161279/C
 Sequence 161279, Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001307
 CURRENT APPLICATION NUMBER: US/09/949, 016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO: 161279
 LENGTH: 601
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-161279

	Query Match	Similarity	1.8%	Score	42.6	DB	4	Length	601	
	Best Local	Similarity	54.0%	Pred. No.	0.22					
	Matches	87	Conservative	0	Mismatches	74	Indels	0	Gaps	0
Qy	2218	AGCTCCCTCACTACGAAAAAGTCACAAATAAATGCAACAGATGTTCTTAATTTGTT						2277		
Db	258	AGTTTCTTCCACCCCTCAAAATTAACAACAAACCAATTACTGGAAAAATTAACACTTC						199		
Qy	2278	TTTCTCTGCTGATGTAATGATTATTATTATATTTTTTTTTTTAGCCGTGATGTGATGA						2337		
Db	198	CTATGGATTTGACCTTAATTTTCTTCATGTCCTTACCTTTACAGSTTAATATATAGGA						139		

OY	2338 AAAGAAAGTTAGAGCCACGATAATCTGAAGCTCACTATT	2378
Db	138 AAAGAAGCTTGACGTCAATGACAATTTGAAGCTGCACAATT	98

```

RESULT 14
US-09-949-016-161280/c
; Sequence 161280, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949.016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,766
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 161280
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-161280

```

	Query Match	Similarity	1.8*	Score	42.6	DB	4	Length	601
	Best Local	Similarity	54.0*	Pred.No.	0.22	74	Indels	0	Gaps
	Matches	87	Conservative	0	Mismatches				
Qy	2218	AGCTCCCTCAACTACAGAAAAGTCACAAATTAATGCAAAATGATGTTCTATTTGTT	2277	TTT	AGCGCGTGAATGTGAT	2337			
Db	554	AGTTTCTTCCACCCCTCAMAATAACAAACAAACATTAAGTGGAAAAATAAACACTTC	495						
Qy	2278	TTTCTGCTGATGATTAAATTGATTATTTATTTT	2277	TTT	TTT	TTT	TTT	TTT	TTT
Db	494	CTATGGATTTGACTTTATTTTCTTCATGTGCTTACCTTTACAGGGTTAATAATATGGA	435						
Qy	2338	AAAGAAAGTTAGAGCCACGATATATCTGAAGCTCACTATT	2378						
Db	434	AAAGGAAGCTTGACGCTCATACATTTTGAAGCTGACAAAT	394						

```

RESULT 15
US-09-385-982-376/C
; Sequence 376, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: 11
; PILE REFERENCE: CCDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 376
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)

```


OTHER INFORMATION: n = A,T,C or G
US-09-385-982-376

Query Match 1.8%; Score 42.6; DB 3; Length 611;
Best Local Similarity 55.0%; Pred. No. 0.23;
Matches 72; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY	2220	CTCCCTCAACTAGAAAAAGTCACATATAAATGCAACATGATGTTCTATTGTTT	2279
Db	351	CNCNTTATATTCCNGTTTACCNGAAGAAAAAGCGNACNTTTTTTTTTTTTTT	292
QY	2280	TCTCTGCTGATGTTAATTATTATTATTATTATTATTATTATTATTATTATT	2339
Db	291	TTTTTTTGNTTTTTTTTTTTTTTTTTTTTTTTATAAGATTATTAGTATATAA	232
QY	2340	AGAAAGTTAGG	2350
Db	231	GGGAGATTAGG	221

Search completed: November 21, 2005, 02:06:35
Job time : 419 secs

Large Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2005, 00:03:25 ; Search time 1699 Seconds

(without alignments)
11589.429 Million cell updates/sec

Title: US-10-099-663-1

Perfect score: 2381
Sequence: 1 agcttcctgcgcgaagaag.....atctgaagctcactatccag 2381

Scoring table: IDENTITY_NTC

Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
23: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
25: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq:*
26: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq:*
27: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
28: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2381	100.0	2381	US-10-099-663-1	Sequence 1, Appl1
2	336	14.1	336	US-10-099-663-2	Sequence 2, Appl1
3	50.6	2.1	16914	US-10-741-601-5698	Sequence 5698, Ap
4	50.6	2.1	16914	US-10-741-600-17777	Sequence 17777, A
5	49.8	2.1	630	US-10-027-632-5694	Sequence 5694, Ap

C	6	49.8	2.1	630	18	US-10-027-632-5694	Sequence 5694, Ap
C	7	49.8	2.1	5204	18	US-10-116-275-303	Sequence 303, App
C	8	48	2.0	546	13	US-09-925-065A-51803	Sequence 51803, A
C	9	48	2.0	569	13	US-09-925-065A-200196	Sequence 200196, A
C	10	48	2.0	570	13	US-09-925-065A-200197	Sequence 200197, A
C	11	48	2.0	814	13	US-09-925-065A-63272	Sequence 63272, A
C	12	47.2	2.0	337	9	US-09-960-352-6036	Sequence 6036, Ap
C	13	47	2.0	2636	26	US-11-097-143-21061	Sequence 21061, A
C	14	46.8	2.0	662	21	US-10-335-053-36	Sequence 36, Appl
C	15	46.8	2.0	670	22	US-10-764-420-2245	Sequence 2245, Ap
C	16	46.2	1.9	6222	16	US-10-311-455-665	Sequence 665, App
C	17	44.4	1.9	565	13	US-09-925-065A-110075	Sequence 110075, A
C	18	44.2	1.9	545	10	US-09-871-161-474	Sequence 474, App
C	19	44.2	1.9	2252	20	US-10-741-601-261	Sequence 261, App
C	20	44.2	1.9	2252	20	US-10-741-601-261	Sequence 745, App
C	21	44.2	1.9	3673778	17	US-10-312-841-2	Sequence 2, Appl1
C	22	43.6	1.8	400	21	US-10-425-115-11538	Sequence 11538, A
C	23	43.6	1.8	413	20	US-10-674-124A-3159	Sequence 2159, Ap
C	24	43.6	1.8	14023	16	US-10-311-455-2077	Sequence 2077, Ap
C	25	43.2	1.8	7143	16	US-10-311-455-956	Sequence 956, Appl
C	26	42.8	1.8	558	13	US-09-925-065A-572321	Sequence 572321, A
C	27	42.8	1.8	679	13	US-09-925-065A-875384	Sequence 875384, A
C	28	42.8	1.8	679	13	US-09-925-065A-875385	Sequence 875385, A
C	29	42.8	1.8	679	13	US-09-925-065A-875386	Sequence 875386, A
C	30	42.8	1.8	51289	20	US-10-322-281-648	Sequence 648, App
C	31	42.6	1.8	513	17	US-10-029-386-13006	Sequence 13006, A
C	32	42.6	1.8	611	10	US-09-871-161-376	Sequence 376, App
C	33	42.6	1.8	1241	21	US-10-425-115-124424	Sequence 124424, A
C	34	42.6	1.8	2844	15	US-10-176-847-19	Sequence 19, Appl
C	35	42.6	1.8	5501	16	US-09-311-455-28	Sequence 28, Appl
C	36	42.4	1.8	452	9	US-09-922-217-138	Sequence 138, App
C	37	42.4	1.8	452	9	US-09-833-263-138	Sequence 138, App
C	38	42.4	1.8	452	14	US-10-025-380-138	Sequence 804505, A
C	39	42.4	1.8	479	13	US-09-925-065A-806507	Sequence 806507, A
C	40	42.4	1.8	481	13	US-09-925-065A-806507	Sequence 804504, A
C	41	42.4	1.8	641	13	US-09-925-065A-804504	Sequence 1721, Ap
C	42	42.4	1.8	682	15	US-10-198-848-1721	Sequence 94, Appl
C	43	42.4	1.8	5908	15	US-10-239-676-94	Sequence 106, App
C	44	42.4	1.8	5908	16	US-10-240-453-106	Sequence 176, App
C	45	42.4	1.8	5908	18	US-10-221-613-176	

ALIGNMENTS

RESULT 1
US-10-099-663-1
; Sequence 1, Application US/10099663
; Publication No. US20030177516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian Gut-Specific Promoters
; FILE REFERENCE: A181
; CURRENT APPLICATION NUMBER: US/10/099,663
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 2381
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(1626)
; NAME/KEY: exon
; LOCATION: (1627)..(1693)
; NAME/KEY: Intron
; LOCATION: (1694)..(2322)
; NAME/KEY: exon
; LOCATION: (2323)..(2381)
; US-10-099-663-1
Query Match 100.0%; Score 2381; DB 17; Length 2381;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 AGCTTCCTGCGCAAAAAGGCTGTGGGGTTCTTTTCCCTCACACAGCTTAAAGCAAAATC 60
Db 1 AGCTTCCTGCGCAAAAAGGCTGTGGGGTTCTTTTCCCTCACACAGCTTAAAGCAAAATC 60
QY 61 CCCAAGTTCAAAACGTCGGCTGTGAAGAGATGCTCACTTCAATGSAAGTAATTATG 120
Db 61 CCCAAGTTCAAAACGTCGGCTGTGAAGAGATGCTCACTTCAATGSAAGTAATTATG 120
QY 121 AAATAATCATAAAACGAGCTCTGTTGGCAGATCAGATTAACCTCTGTGGGCAAAAT 180
Db 121 AAATAATCATAAAACGAGCTCTGTTGGCAGATCAGATTAACCTCTGTGGGCAAAAT 180
QY 181 CTTAAAGTGTATGTGTAGAAACAGAGAGTGTGGCACTAAATGACATGATTTGACACA 240
Db 181 CTTAAAGTGTATGTGTAGAAACAGAGAGTGTGGCACTAAATGACATGATTTGACACA 240
QY 241 TTTGATCTTCTAGGAGCAAAAAGGGTCTGGAAAAACAATTAATTCGTGTGCAAGTCACT 300
Db 241 TTTGATCTTCTAGGAGCAAAAAGGGTCTGGAAAAACAATTAATTCGTGTGCAAGTCACT 300
QY 301 AGCAGCCTGTTTTGGGTGCACTACAGCAACTTTGTTGGCAACAATTAACATCTAAGTTG 360
Db 301 AGCAGCCTGTTTTGGGTGCACTACAGCAACTTTGTTGGCAACAATTAACATCTAAGTTG 360
QY 361 TTTTCTTTTCTCTTCTTCTTAACTTCTGTACAGTCTAAAGTGTAAAGTGTATGTAAGT 420
Db 361 TTTTCTTTTCTCTTCTTCTTAACTTCTGTACAGTCTAAAGTGTAAAGTGTATGTAAGT 420
QY 421 TACTTCCCTCTGTCAGTCTCTTACGCAAGTATGAGCATTTGCAAAAATGAACTGAGTGA 480
Db 421 TACTTCCCTCTGTCAGTCTCTTACGCAAGTATGAGCATTTGCAAAAATGAACTGAGTGA 480
QY 481 ATGGAAGGCCACCTATTTTGTATCAACAGCAAAAGTTCTAAATTTATAGTTATACCTTCA 540
Db 481 ATGGAAGGCCACCTATTTTGTATCAACAGCAAAAGTTCTAAATTTATAGTTATACCTTCA 540
QY 541 GTTAAACCTTTTCTGTCAGGCTGTGAAAGAAAAGAAATTATGATTAACACAGACTAGTA 600
Db 541 GTTAAACCTTTTCTGTCAGGCTGTGAAAGAAAAGAAATTATGATTAACACAGACTAGTA 600
QY 601 AAATTCATTAGTTAGAGCCAAACCTGTATCTGTGTATGTAAGCAACATTCATTGAGCAT 660
Db 601 AAATTCATTAGTTAGAGCCAAACCTGTATCTGTGTATGTAAGCAACATTCATTGAGCAT 660
QY 661 TCAGATTTACATTTTGAAGCTAATAGACAGCATTTGGTCCGTCCATAGGAACAGA 720
Db 661 TCAGATTTACATTTTGAAGCTAATAGACAGCATTTGGTCCGTCCATAGGAACAGA 720
QY 721 CTAACTATTAATCTGTAGATTAGTACAAAGCAATTTAGCACAGCAAAATTTGCTCAAGTTTC 780
Db 721 CTAACTATTAATCTGTAGATTAGTACAAAGCAATTTAGCACAGCAAAATTTGCTCAAGTTTC 780
QY 781 AAGTAGACATATCTGTGGGGAAGAGAGCTGAGGCAAGTGTGCTCAATTTCTGCAAT 840
Db 781 AAGTAGACATATCTGTGGGGAAGAGAGCTGAGGCAAGTGTGCTCAATTTCTGCAAT 840
QY 841 ATCTTCACACATTTAAACCTGGGATCTATGAAATCAACACGTTGGGTAAATTCAT 900
Db 841 ATCTTCACACATTTAAACCTGGGATCTATGAAATCAACACGTTGGGTAAATTCAT 900
QY 901 TAGCAGACATCAACTCTGTAGGAATGACAGAAACAGAGCATTCATCTAATGAGGCTAT 960
Db 901 TAGCAGACATCAACTCTGTAGGAATGACAGAAACAGAGCATTCATCTAATGAGGCTAT 960
QY 961 AATATAGAGATCGTAGAGGTGCTGTAATTTAGACTAATTAAGAGTGAAGGACA 1020
Db 961 AATATAGAGATCGTAGAGGTGCTGTAATTTAGACTAATTAAGAGTGAAGGACA 1020
QY 1021 CGAATGAGAAATATCATCGCAATTTCTGTAGCTCAGACCTAGACTCGAAGGTTTCTGAAA 1080
Db 1021 CGAATGAGAAATATCATCGCAATTTCTGTAGCTCAGACCTAGACTCGAAGGTTTCTGAAA 1080
```

```
Db 1021 CGAATGAGAAATATCATCGCAATTTCTGTAGCTCAGACCTAGACTCGAAGGTTTCTGAAA 1080
QY 1081 CTGAACCGAAGTTTCCCAAACTTACCTGTGGAATGTTTACGTGATTCCTTCAATCTCTAT 1140
Db 1081 CTGAACCGAAGTTTCCCAAACTTACCTGTGGAATGTTTACGTGATTCCTTCAATCTCTAT 1140
QY 1141 TATGTGAGTGAATATGATTTCTACCAAAATTAAGATGACAAAGCAGAGATTTGTGTTTT 1200
Db 1141 TATGTGAGTGAATATGATTTCTACCAAAATTAAGATGACAAAGCAGAGATTTGTGTTTT 1200
QY 1201 ATCTGTGGGTAAATAGCGTTTCTCCAGTGTATAAAGACCTCCACACAGTATTAAGTC 1260
Db 1201 ATCTGTGGGTAAATAGCGTTTCTCCAGTGTATAAAGACCTCCACACAGTATTAAGTC 1260
QY 1261 CTATGCAACAAAGAAATGTCATATCTCTTATGCTCATTAATTAATTTTCAATAGATA 1320
Db 1261 CTATGCAACAAAGAAATGTCATATCTCTTATGCTCATTAATTAATTTTCAATAGATA 1320
QY 1321 GCCGGTTTTTACTCAACTCAAAATAGATGAAACGAATGAAATGGGTTAGTACTGTTTA 1380
Db 1321 GCCGGTTTTTACTCAACTCAAAATAGATGAAACGAATGAAATGGGTTAGTACTGTTTA 1380
QY 1381 TAAAGAAAGTAAATAAAGATACTATCATTTTGAAGCAATAGGAGAGAGATTCAG 1440
Db 1381 TAAAGAAAGTAAATAAAGATACTATCATTTTGAAGCAATAGGAGAGAGATTCAG 1440
QY 1441 CAAGAAGTGTGCTTCAAGTGGAAAACAAGTAACTAAAGTACCCCTCTTGACAA 1500
Db 1441 CAAGAAGTGTGCTTCAAGTGGAAAACAAGTAACTAAAGTACCCCTCTTGACAA 1500
QY 1501 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTGATA 1560
Db 1501 GATCAATGCCACAGTTGAGCTTTAGCCAGCCACATCATCATGTAATTTGCTTCTGATA 1560
QY 1561 AGCTGTCTATTAATTTCTTTTGAAGCTGTCTACTTACCAAGATGTGCTTACAGAC 1620
Db 1561 AGCTGTCTATTAATTTCTTTTGAAGCTGTCTACTTACCAAGATGTGCTTACAGAC 1620
QY 1621 AGAAGATGGCATTTTAAAGGTCTTGGAAAATAGGAAAATGAGAACATGTAAGAAAATTC 1680
Db 1621 AGAAGATGGCATTTTAAAGGTCTTGGAAAATAGGAAAATGAGAACATGTAAGAAAATTC 1680
QY 1681 ATGGAAGCAATGGTAAGCTTACTTTTGTGAATGCTCTTCTAAAGCAGAGATACACATAC 1740
Db 1681 ATGGAAGCAATGGTAAGCTTACTTTTGTGAATGCTCTTCTTAAAGCAGAGATACACATAC 1740
QY 1741 GCGGGAATACAAACTTAAGCTGTACAGAACATCAATCTGCTAACTCTGCTTGTGT 1800
Db 1741 GCGGGAATACAAACTTAAGCTGTACAGAACATCAATCTGCTAACTCTGCTTGTGT 1800
QY 1801 CTGCTATTTTGGCCCTTGCACATTTGCCCTGCATTAATTTGAAAAGCTCTATAGAGGGGA 1860
Db 1801 CTGCTATTTTGGCCCTTGCACATTTGCCCTGCATTAATTTGAAAAGCTCTATAGAGGGGA 1860
QY 1861 ATACAAGAAAGAAAACATTCGATTTTATTTGCAATTCGATATCTTATGATTTAGCT 1920
Db 1861 ATACAAGAAAGAAAACATTCGATTTTATTTGCAATTCGATATCTTATGATTTAGCT 1920
QY 1921 AATTCAGTAGAGGCAATTCACAGAAATTTAATGAAATTAATGTAATGAAATATATAT 1980
Db 1921 AATTCAGTAGAGGCAATTCACAGAAATTTAATGAAATTAATGTAATGAAATATATAT 1980
QY 1981 TTGATAGACGTGTTGAAAATTTACCAAGAGGAAATGCGGTGCTCAGTTTGGCAGA 2040
Db 1981 TTGATAGACGTGTTGAAAATTTACCAAGAGGAAATGCGGTGCTCAGTTTGGCAGA 2040
QY 2041 ACACACATGATTTGAGTCATTTTAAACATGCTAGTCTTAACTTTAAAGCTTTGTA 2100
Db 2041 ACACACATGATTTGAGTCATTTTAAACATGCTAGTCTTAACTTTAAAGCTTTGTA 2100
QY 2101 CTGTAATATGATGATTAACATTAATCTTCTAGTTGATAGTGTGTAATTAACAGGCTGA 2160
Db 2101 CTGTAATATGATGATTAACATTAATCTTCTAGTTGATAGTGTGTAATTAACAGGCTGA 2160
```

Qy	2161	ACACTGCTCAGTGAAGGTGGAAGAGTAACTCTGAGTCAGAAATCTGGGCTAAGC	2220
Db	2161	ACACTGCTCAGTGAAGGTGGAAGAGTAACTCTGAGTCAGAAATCTGGGCTAAGC	2220
Qy	2221	TCCTCACTACAGAAAAAGTCACATTAATAATGCAACATGATGTTCTATTTGTTTT	2280
Db	2221	TCCTCACTACAGAAAAAGTCACATTAATAATGCAACATGATGTTCTATTTGTTTT	2280
Qy	2281	CTCTGCTTGATGTTAATTGATTATTTTTTTTTTAAAGCGTGATGTATGAAAA	2340
Db	2281	CTCTGCTTGATGTTAATTGATTATTTTTTTTTTAAAGCGTGATGTATGAAAA	2340
Qy	2341	GAAAGTTAGAGCCACGATTAATCGAAGCTCACATTTCAAG	2381
Db	2341	GAAAGTTAGAGCCACGATTAATCGAAGCTCACATTTCAAG	2381

```

RESULT 2
US-10-099-663-2
; Sequence 2, Application US/10099663
; Publication No. US20030177516A1
; GENERAL INFORMATION:
; APPLICANT: Avigenics, Inc
; TITLE OF INVENTION: Avian GUT-Specific Promoters
; FILE REFERENCE: A181
; CURRENT APPLICATION NUMBER: US/10/099,663
; CURRENT FILING DATE: 2002-03-14
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Gallus gallus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(336)
; OTHER INFORMATION: chick intestinal fatty acid binding protein promoter region
; US-10-099-663-2

```

Query Match	14.1%	Score 336	DB 17	Length 336
Best Local Similarity	100.0%	Pred. No. 5.4e-74		
Matches 336	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1301	ATTATTATTTTCATTAGATAGCCGGTTTTTTACTACAACTCAATTAAGATGAACAGATG	1360	
Db	1	ATTATTATTTTCATTAGATAGCCGGTTTTTTACTACAACTCAATTAAGATGAACAGATG	60	
QY	1361	AATGGGTTAGTGCATGCTTTTATTAAGAAAGATTAATTAAGATACATATCATATTGAGGCAA	1420	
Db	61	AATGGGTTAGTGCATGCTTTTATTAAGAAAGATTAATTAAGATACATATCATATTGAGGCAA	120	
QY	1421	TAAAGGAGGGAGAGATTCACGAAACAGTGGCTTACAAGTGGAGAAACAAGTTAACTTAA	1480	
Db	121	TAAAGGAGGGAGAGATTCACGAAACAGTGGCTTACAAGTGGAGAAACAAGTTAACTTAA	180	
QY	1481	GTGACCCCCCTCCTCTTGACAAGATCAATGGCCACAGTTAGCTTTGAGCCAGCCACATCA	1540	
Db	181	GTGACCCCCCTCCTCTTGACAAGATCAATGGCCAGTTAGCTTTGAGCCAGCCACATCA	240	
QY	1541	TGTAAATTTGCTTTCCTGATAAGCCTGTTTCATTAATTTCTTTTGGAAAAGCTTCGTACTTA	1600	
Db	241	TGTAAATTTGCTTTCCTGATAAGCCTGTTTCATTAATTTCTTTTGGAAAAGCTTCGTACTTA	300	
QY	1601	CCAGAACTCTGCCCTACAGACAGAAAAGATGGCATTTA	1636	
Db	301	CCAGAACTCTGCCCTACAGACAGAAAAGATGGCATTTA	336	

RESULT 3
US-10-741-601-5698
; Sequence 5698, Application US/10741601
; Publication No. US20040166519A1

```

/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: C1001500
/ CURRENT APPLICATION NUMBER: US/10/741,601
/ CURRENT FILING DATE: 2003-12-22
/ NUMBER OF SEQ ID NOS: 26415
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5698
/ LENGTH: 16914
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-10-741-601-5698

```

Query March	2.1*	Score 50.6;	DB 20;	Length 16914;
Best Local Similarity	58.8*	Pred. No. 0.46;		
Matches 124; Conservative	2;	Mismatches	76;	Indels 9;
				Gaps 2;

QY	1516	TGAGCTTTAGCCAGCCAGCATCATCATTAATGCTTCCGAAATAGCCTGTCATAAT	1575
QY	1516		
Db	5945	TGAACCTTTAGCCTTCCACATCACTATATGAAGTTGGTTAAGATTAAGATAATATTAAT	6004
QY	1576	TCTC-----TTTGCAAGCTCTGCTACTTACCAAGGCTGCTACAGACAGAAAGA	1627
Db	6005	TCTTGCCCAAGAGACAGACTGAACTCTTAAGCTGCTAGAGGCTGACT-CAACTGAAATCA	6066
QY	1628	TGCGATTAAAGGTACTTGGAAAAATAGAAAAATGGAACCTATGAAAAATTCATGGAAG	1687
Db	6064	TGGCGTTTGACACACACTTGGAAAGGTAGACCGGAGATGAAAACTATGACAAAGTTATGGA	6123
QY	1688	CAATGGGTAAAGCCTTACTTTTGGATGCGCT	1718
Db	6124	AAATGGGTAAAGACTTTATTTCTTTGGGCT	6154

```

RESULT 4
US-10-741-600-17777
; Sequence 17777, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ. ID NOS: 73997
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ. ID NO 17777
; LENGTH: 16914
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
; US-10-741-600-17777

```

Query Match	2.1 %	Score	50.6	DB	22	Length	16914
Similarity	58.8 %	Pred. NO.	0.46				
Best Local		Matches	124	Conservative	2	Mismatches	76
						Indels	9
						Gaps	2

[illegible]

Db 6124 AATGGGTAAAGACTTATTCTTTGTGGCT 6154

RESULT 5

US-10-027-632-5694/c
Sequence 5694, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 1999-11-23

PRIOR FILING DATE: 1999-09-28

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 5694

LENGTH: 630

TYPE: DNA

ORGANISM: Human

US-10-027-632-5694

Query Match

Best Local Similarity 59.2%; Pred. No. 0.12; Length 630;

Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;

Db 1516 TGAGCTTTAGCCAGCATCATCATGTAATGCTTCTGTAAGCTGTTCAAT 1575

Db 531 TGAACCTTTAGCTTCCACATCATCATGTAAGCTTCTGTAAGCTGTTCAAT 472

QY 1576 TCTC-----TTGCAAGCTCTGCTACTTACCAAGTCTGCTTACGACAGAA 1627

Db 471 TCTGCCCAAGACAGACCTGTAATCTTACCTGCTTACGAGCTGACT-CAACTGAATCA 413

QY 1628 TGACATTTAAGCGTACTTGGAAATAGAAATAGAAATAGAAATAGAAATAGAA 1687

Db 412 TGCGCTTTGACAGACCTTGGAAATAGAAATAGAAATAGAAATAGAAATAGAA 353

QY 1688 CAATGGGTAGCCTTACTTTTGAATGCT 1718

Db 352 AATGGGTAAAGACTTATTCTTTGTGGCT 322

RESULT 6

US-10-027-632-5694/c

Sequence 5694, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 5694

LENGTH: 630

TYPE: DNA

ORGANISM: Human

US-10-027-632-5694

Query Match

Best Local Similarity 59.2%; Pred. No. 0.12; Length 630;

Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;

QY 1516 TGAGCTTTAGCCAGCATCATCATGTAATGCTTCTGTAAGCTGTTCAAT 1575

Db 531 TGAACCTTTAGCTTCCACATCATCATGTAAGCTTCTGTAAGCTGTTCAAT 472

QY 1576 TCTC-----TTGCAAGCTCTGCTACTTACCAAGTCTGCTTACGACAGAA 1627

Db 471 TCTGCCCAAGACAGACCTGTAATCTTACCTGCTTACGAGCTGACT-CAACTGAATCA 413

QY 1628 TGACATTTAAGCGTACTTGGAAATAGAAATAGAAATAGAAATAGAAATAGAA 1687

Db 412 TGCGCTTTGACAGACCTTGGAAATAGAAATAGAAATAGAAATAGAAATAGAA 353

QY 1688 CAATGGGTAGCCTTACTTTTGAATGCT 1718

Db 352 AATGGGTAAAGACTTATTCTTTGTGGCT 322

RESULT 7

US-10-116-275-303

Sequence 303, Application US/10116275

Publication No. US20030211476A1

GENERAL INFORMATION:

APPLICANT: Eian Pharmaceutical Technology

APPLICANT: O'Mahony, Daniel J.

APPLICANT: Brayden, David

APPLICANT: Byrne, Dargah

APPLICANT: Lambkin, Imelda

APPLICANT: Higgins, Lisa

TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and

FILE REFERENCE: E1067/20087

CURRENT APPLICATION NUMBER: US/10/116,275

CURRENT FILING DATE: 2002-10-04

NUMBER OF SEQ ID NOS: 349

SOFTWARE: Patencin version 3.1

SEQ ID NO 303

LENGTH: 5204

TYPE: DNA

ORGANISM: Homo sapiens

US-10-116-275-303

Query Match

Best Local Similarity 59.2%; Pred. No. 0.38; Length 5204;

Matches 125; Conservative 0; Mismatches 77; Indels 9; Gaps 2;

QY 1516 TGAGCTTTAGCCAGCATCATCATGTAATGCTTCTGTAAGCTGTTCAAT 1575

Db 971 TGAACCTTTAGCTTCCACATCATCATGTAAGCTTCTGTAAGCTGTTCAAT 1030

QY 1576 TCTC-----TTGCAAGCTCTGCTACTTACCAAGTCTGCTTACGACAGAA 1627

Db 1031 TCTGCCCAAGACAGACCTGTAATCTTACCTGCTTACGAGCTGACT-CAACTGAATCA 1089

[illegible]

; ORGANISM: Mus musculus
US-10-335-053-36

Query Match
Best Local Similarity 2.0%; Score 46.8; DB 21; Length 662;
Best Local Similarity 73.2%; Pred. No. 0.7;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1615 ACAGACGAGAAAGATGGCATTTAAGGTACTTGGAAATAGAGAAATGAGAACTATGAA 1674
|||
Db 28 ACAGCTGAGATCATGGCATTCGACGCGACGTGGAAAGTAGACCGAAGAGAACTATGAA 87
|||

QY 1675 AAATTCATGGAGCAATGGGTA 1696
|||
Db 88 AAGTTCATGGAGAAATGGGCA 109
|||

RESULT 15
US-10-764-420-2245
; Sequence 2245, Application US/10764420
; Publication No. US20050084872A1
; GENERAL INFORMATION:
; APPLICANT: Lum, Pek Yee
; APPLICANT: Tan, Yejun
; APPLICANT: Dai, Hongyue
; TITLE OF INVENTION: Methods For Determining Whether An Agent
; TITLE OF INVENTION: Possesses A Defined Biological Activity
; FILE REFERENCE: ROSA12057
; CURRENT APPLICATION NUMBER: US/10/764,420
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/442,797
; PRIOR FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US 60/474,413
; PRIOR FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 3683
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2245
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-764-420-2245

Query Match
Best Local Similarity 2.0%; Score 46.8; DB 22; Length 670;
Best Local Similarity 73.2%; Pred. No. 0.7;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1615 ACAGACGAGAAAGATGGCATTTAAGGTACTTGGAAATAGAGAAATGAGAACTATGAA 1674
|||
Db 24 ACAGCTGAGATCATGGCATTCGACGCGACGTGGAAAGTAGACCGAAGAGAACTATGAA 83
|||

QY 1675 AAATTCATGGAGCAATGGGTA 1696
|||
Db 84 AAGTTCATGGAGAAATGGGCA 105
|||

Search completed: November 21, 2005, 05:42:20
Job time : 1705 secs

100-100000 (uspio)